BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

REMARKS

Claims 1 to 3 and 6 to 17 are pending. Claims 18 and 19 have been added. Thus, claims 1 to 3 and 6 to 19 are presently under examination.

Regarding the claim amendments and new claims

Claims 8 and 16 have been amended to indicate that the recited library of diverse molecules is a library of diverse peptides or peptidomimetics. The amendment to claims 8 and 16 is supported throughout the specification as originally filed, for example, at page 6, lines 24-31, which indicates that a molecule can be a peptide including a variant or modified peptide or peptide-like molecule such as a peptidomimetic.

New claims 18 and 19 have been added. New claim 18 is directed to a method of identifying a molecule that homes to a selected organ or tissue, where identification of the homing molecule includes mass spectrometry. New claim 18 is supported throughout the specification, for example, at page 8, lines 11-15, which indicates that a highly sensitive detection method such as mass spectrometry, either alone or in combination with a method such as gas chromatography, can be used to identify organ homing molecules in a sample of a selected organ or tissue.

New claim 19 is directed to a method of identifying a molecule that homes to a selected organ or tissue by administering to a subject a library of diverse molecules other than a nucleic acid library; collecting a sample of the selected organ or tissue; identifying a molecule that homes to the selected organ or tissue; and individually administering to a subject the identified molecule which homes to the selected organ or tissue. New claim 19 is supported throughout the specification, for example, at page 36, lines 15-27, which discloses that, following identification of several brain-homing peptide motifs, phage displaying the predominant motifs were amplified individually, administered to mice, and recovered from brain and from a control organ, kidney.

As set forth above, the amendments and new claims are supported in the specification as originally filed and do not add new matter. Applicants therefore respectfully request that the Examiner enter the amendments and new claims.

Regarding the Examiner interview

Applicants appreciate the Examiner's consideration and discussion of the remaining issues in the case in the personal interview with Dr. Erkki Ruoslahti and Ms. Cathryn Campbell on April 20, 2004. The enablement rejection was discussed in view of a new Rule 132 Declaration with new mass spectrometric data corroborating that the claimed methods can be practiced as set forth in the specification. The Examiner indicated the sufficiency of these corroborating data to remove the outstanding enablement rejection. The amendment of claims 8 and 16 was further discussed in regard to the outstanding written description rejection.

Regarding the terminal disclaimer

As acknowledged on page 2 of the Office Action mailed November 26, 2003, a terminal disclaimer was filed on September 29, 2003, disclaiming the terminal portion of any patent granted on the subject application which would extend beyond the expiration date of U.S. Patent Nos. 5,622,699; 6,068,829; 6,296,832; or 6,306,365. Applicant appreciates the Examiner's acknowledgement that the terminal disclaimer has been reviewed and accepted and, further, has been recorded.

Regarding the written description rejection

The rejection of claims 8 and 16 under 35 U.S.C. § 112, first paragraph, as allegedly lacking written description, is respectfully traversed.

Claims 8 and 16 are directed to methods practiced with "a library of diverse peptides and peptidomimetics." Although the Examiner acknowledges that the specification supports that the molecules of a library can be "peptides or peptide-like molecules," it is alleged

that written description is provided in the specification for a library which includes a mixture of peptides and peptidomimetics.

Applicants maintain that the specification provides written description for the libraries of claims 8 and 16, which are made up of diverse "peptides and peptidomimetics." Firstly, the specification indicates that a library is a "collection of molecules" and does not indicate that the molecules must be of a uniform type. Secondly, as is explicitly set forth in the specification as originally filed, the term "peptide" is used in the subject application to include peptides, proteins, peptidomimetics and the like. Thus, in view of the lexicography of the subject application, a library of "diverse peptides" can include both peptides and peptidomimetics. Although Applicants maintain that it is clear to the skilled person that Applicants were in possession of the claimed invention including the invention as it would be practiced with a diverse library of peptides and peptidomimetics, claims 8 and 16 have been amended herein in order to further prosecution of the subject application without prejudice to Applicants' pursuing the subject matter of the original claims in a continuation application claiming the benefit of priority of the subject application.

In view of the above remarks and amendments, the Examiner is respectfully requested to reconsider and remove the rejection of claims 8 and 16 under 35 U.S.C. § 112, first paragraph, as allegedly lacking written description.

Regarding the enablement rejection

The rejection of claims 1 to 3 and 6 to 17 under 35 U.S.C. § 112, first paragraph, as allegedly lacking enablement for the full scope of the invention, is respectfully traversed. Each of the grounds for rejecting the claims as allegedly lacking enablement is addressed in turn below.

Regarding the use of recovered molecules

The previous Office Action indicates that the specification teaches but a single use for a molecule that homes to a selected organ: identification of the library member so that the homing molecule can be produced and used as a targeting ligand. In asserting that only identified end products have utility, the Office Action takes the position that the claims must be enabled for the further step of identifying individual library members. The Office Action acknowledges that the specification discloses several means (tags) for recovery of a homing molecule, including a physical support such as a microbead or phage, and biotin. However, it is alleged that the specification teaches but a single type of tag for identification of an individual library member, a nucleic acid molecule.

In parent application 09/227,906, Applicants have argued, and the Examiner has acknowledged, that one would have been able to use recovered homing molecules without knowing their structure. Applicants herein make of record in the present case the arguments presented in the response to Office Action filed April 6, 2000, in parent application Serial No. 09/227/906. Specifically, Applicants maintain that identification of a recovered homing molecule is not necessary to practice the invention of claims 1 to 3 and 6 to 9. Applicants respectfully disagree that the only utility for the claimed invention is as an intermediate in a process for identifying the homing molecule.

Specifically, a recovered "molecule that homes to a selected organ or tissue" is itself clearly useful as a binding reagent that can be linked, for example, to a detectable moiety such as a radiolabel for diagnostic purposes. In the same manner that a monoclonal antibody is valuable even when its structure has not been determined, a recovered homing molecule of the invention can be used as a specific binding agent even without knowledge of the structure of the molecule. Given that the organ homing specificity of the recovered molecule, for example, brain-homing specificity or kidney-homing specificity will be known, one skilled in the art

would have had sufficient knowledge of the homing molecule to use it as a binding reagent, for example, to identify an organ sample.

In sum, Applicants maintain that a recovered "molecule that homes to said selected organ or tissue" has utility as a binding reagent with a known specificity, even without "identification" of the molecule or knowledge of its structure. In view of the above remarks, Applicants respectfully request that the Examiner remove this ground for rejecting claims 1 to 3 and 6 to 17 under 35 U.S.C. § 112, first paragraph, as allegedly lacking enablement.

Regarding selection of starting libraries

As a further ground for rejecting the claims as allegedly lacking enablement, the Office Action asserts that the specification does not teach any and all possible libraries of organic compounds nor which libraries would be likely to yield homing molecules. Applicants have previously argued in parent application 09/227,906, and the Examiner has acknowledged, that one of skill in the art would have been able to prepare various types of libraries. These arguments and evidence are made of record below.

There are two issues which support this ground for rejection. Firstly, the Office Action indicates that the specification does not teach how to make any and all libraries of organic molecules, although it is acknowledged that techniques for making peptide libraries as well as certain peptidomimetic libraries were known in the art. In this regard, the Office Action alleges that well-known techniques do not extend to any and all possible libraries of organic compounds, or to any and all possible libraries of peptidomimetics. Secondly, the specification allegedly does not suggest which of all possible organic molecule libraries other than polypeptides and peptidomimetics would be likely to yield at least one homing molecule, and further asserts that this alleged deficiency is not remedied by what was well known in the art.

1. Construction of any and all possible libraries

As discussed further below, it is not necessary that the specification teach "any and all possible libraries" suitable for practicing the claimed invention. Rather, an invention may be enabled in view of what is well known in the art in combination with the teachings of the specification. In this regard, the claimed methods can be practiced without undue experimentation using, for example, non-peptide libraries including polymeric and non-polymeric libraries such as organic chemical libraries as set forth in the specification. Such a chemical library can be, for example, a library of natural product molecules, or a library of synthetic molecules such as a decorated monomer library or oligomer library. As evidence that non-peptide libraries such as chemical libraries were well known in the art at the time the invention was made, Applicants submit herewith Exhibit 1 by Ecker and Crooke, Biotechnology 13:351-360 (1995), a review article indicating that a variety of chemical libraries were available for drug discovery in 1995. In addition, Gordon et al., J. Medicinal Chemistry 37: 1385-1401 (1994), attached as Exhibit 2, describe construction of several nonpolymeric small molecule libraries including a benzodiazepine library and a hydandoin library (see pages 1390-1392). As further evidence that one skilled in the art would have been able to prepare organic chemical libraries without undue experimentation, Applicants submit as Exhibit 3 a publication by Burbaum et al., Proc. Natl. Acad. Sci., *USA* 92:6027-6031 (1995). Burbaum et al. produce two libraries, a dihydrobenzopyran library and an acylpiperidine library, to be screened for carbonic anhydrase-binding activity (page 6028, first column, last paragraph). The authors explicitly comment that "very large combinatorial libraries of small molecules on solid supports can now be synthesized" (see abstract). In sum, these publications corroborate that a variety of non-peptide organic chemical libraries, including non-polymeric chemical libraries, were known in the art at the time the priority application was filed.

Furthermore, Applicants maintain that it is not required to disclose or provide evidence showing availability of "all possible organic chemical libraries." Rather, all that is required is that sufficient teaching be present in the specification such that, combined with what

is well known in the art, the invention can be practiced without undue experimentation. As indicated by the Examiner in the telephonic interview of August 11, 2000, pertaining to parent application Serial No. 09/227/906, the *in vivo* panning invention is a generic invention. As further acknowledged by the Examiner, no inventor can be expected to describe or be aware of all species of libraries encompassed by the claims.

In sum, Applicants have provided evidence of the availability of a variety of non-peptide organic chemical libraries including non-polymeric organic chemical libraries. Furthermore, in parent application Serial No. 09/227/906, the same issue was overcome. See, for example, Paper #10, Office Action dated December 14, 2000, which indicates that "Applicant's arguments concerning the breadth of the 'diverse molecules'...are convincing" (page 5, second sentence). In view of what is taught in the specification and the corroborating evidence presented herein exemplifying that a variety of libraries were known in the art, Applicants respectfully request that this ground for rejection be removed.

2. Selection of libraries which will yield homing molecules

Applicants would make several additional points in response to the assertion that, since the claimed invention borders on a pioneering invention, it cannot have been routine in the art to make and test various types of libraries to determine the identity of types of libraries that would provide at least one member that would home to a selected organ or tissue.

a. Firstly, in the methods of the invention, libraries of diverse molecules are screened based on the general expectation, borne out by practice as described further below, that a random library which is sufficiently large will contain at least one molecule that homes to the selected organ or tissue. Again, this expectation has been borne out by practice. In this regard, Applicants point to the Rule 132 Declaration executed by Dr. Erkki Ruoslahti as corroboration that one skilled in the art would have been able to identify homing molecules from a random library of diverse molecules without undue experimentation. In particular, the attached

Declaration indicates that several homing molecules were identified from a library of 75 small organic molecules randomly selected from a 420,000 molecule ChemBridge library.

b. Secondly, Applicants submit that it would have been well within the level of one of skill in the art to mix several drug or organic chemical libraries before administering the "mixed" library to a subject, as specified in step (a) of the present methods. Mixing of libraries is taught in the specification, for example, in Example II, which discloses *in vivo* panning in mice where brain is the selected organ. As disclosed in the specification,

> Mice were injected with two different mixtures of phage libraries. The first mixture contained libraries encoding CX_9 (SEQ ID NO: 39), CX_5C (SEQ ID NO: 36), CX_6C (SEQ ID NO: 37) and CX_7C (SEQ ID NO: 38) peptides (CX_{5-7}/CX_9 mixture; SEQ ID NOS: 36-39). The second mixture contained libraries encoding $X_2CX_{14}CX_2$ (SEQ ID NO: 40) and X_2CX_{18} (SEQ ID NO: 41) peptides ($X_2CX_{18}/X_2CX_{14}CX_2$ mixture; SEQ ID NOS: 40 and 41). The phage library mixtures were administered to mice via tail vein injection.
> Specification at page 33, lines 16-26

Applicants submit to the Examiner that it would not have been beyond the level of skill of one of average skill in the art to mix non-peptide libraries in the same manner as phage-displayed peptide libraries were mixed in Example II. Such mixing of libraries in order to create a more complex library to be administered to a subject obviates concerns regarding selecting one particular non-peptidic library to be used in practicing the claimed invention.

c. Thirdly, focused libraries designed around a "lead compound" for optimization of binding and other characteristics were well known in the art at the time the invention was made. Thus, rather than administering a random library to a subject without relying on any information regarding molecules known to selectively home to the selected organ or tissue, one skilled in the art would have been able to use a library of diverse molecules focused around a known homing

molecule or a molecule known to bind to a target receptor expressed in the selected organ or tissue. Where the known homing or binding molecule is a peptide, such a focused library can be a library of diverse non-peptide molecules or a library of diverse peptide molecules. The use of such smaller, focused libraries was well within the level of the ordinary artisan at the time the invention was made. As stated above, one skilled in the art would have been able to prepare or select focused non-peptidic libraries, if desired, based on the structure of a known homing peptide or a peptide known to bind a target receptor expressed in the organ of interest.

d. Fourthly, even if, for the sake of argument, one skilled in the art were to practice steps (a) and (b) with a particular library as set forth in claim 1 and were not to recover or identify a homing molecule, such would not preclude enablement. Rather, if the molecules recovered from a particular library did not contain a homing molecule, the performed procedure would not be encompassed by claim 1. Similarly, if one skilled in the art were to practice steps (a) and (b) with a particular library as set forth in claim 10 and were to fail to identify a homing molecule, such would not preclude enablement. Instead, such a procedure simply would not be encompassed by claim 10. In this regard, it is well established that claims are permitted to embrace one or more inactive embodiments.

In view of the evidence made of record above and further in view of the corroborating evidence presented in the attached Rule 132 Declaration, Applicants submit that undue experimentation would not have been required to select a library of diverse molecules appropriate for identifying a homing molecule using *in vivo* panning. Applicants would also emphasize that the issues of whether libraries would have been prepared by routine methods and whether libraries would yield homing molecules is the same for tagged and untagged libraries. Applicants maintain that one of skill in the art would have been able to select from a variety of libraries a library to be administered in the methods of the invention. In view of the above remarks, and further in view of the corroborating evidence discussed with the Examiner in the recent interview and submitted herein, Applicants respectfully request that the Examiner remove

this ground for rejecting the claims as allegedly lacking enablement under the first paragraph of 35 U.S.C. § 112.

Regarding the Rule 132 Declaration of record

Applicants have previously submitted a Rule 132 Declaration showing *in vivo* panning results obtained with a 10-compound library. The Examiner believes that the Declaration describes a pre-defined library of 10 different benzodiazepines, and asserts that the specification does not teach using pre-defined libraries. Applicants wish to clarify that the library of 10 compounds includes multiple benzodiazepines, but also contains compounds with different structures. Thus, the work described in the Declaration does not rely on a "pre-defined benzodiazepine library."

The Office Action further queries the multiple peaks observed for a single compound, and the inaccuracy of the molecular weights observed. In this regard, Applicants submit that the suboptimal quality of the starting library compounds used is consistent with additional peaks observed. Furthermore, the molecular weight inaccuracies observed likely resulted from mass spectrometric techniques which were not optimal for the size of the library components being analyzed. In particular, the spectrometer used in this experiment was calibrated for protein, rather than small molecule, detection.

Finally, the Examiner notes that Figures 1 and 6 of the previously submitted Declaration are both from brain extracts yet reveal distinct spectra. Applicants submit that both Figures 1 and 6 show mass spectrometric results of brain extracts from mice injected with the 10-compound library. The distinct appearance of the two spectra is explained by the fact that each figure shows different portions of the spectra, with Figure 1 showing m/z from 268 to 310 and Figure 6 showing m/z from 220 to 250.

Applicants maintain that the previously submitted Rule 132 Declaration supports that the invention can be practiced as claimed with untagged small molecule libraries. In

addition, the teachings of the specification have been further corroborated using a random library of 75 commercially prepared small molecules, as evidenced in the Rule 132 Declaration submitted herewith.

Regarding enablement of untagged libraries of peptides

Applicants have previously argued that one skilled in the art would have been able to identify peptides from untagged libraries and have provided Clauser et al. as evidencing that one of skill in the art would have been able to use, for example, HPLC and mass spectrometry to identify untagged peptides from a peptide library. For convenience, a copy of the previously exhibited publication by Clauser et al. is attached hereto as Exhibit 4. The Office Action responds to Applicants' arguments by asserting that the specification does not teach such methods or how to use them in identifying homing peptides from a library.

First, Applicants submit that, in view of what was well known in the art at the time the invention was made, the specification provides sufficient guidance to practice the claimed invention using untagged peptides. In particular, the specification teaches that an organ extract can be separated based on molecular weights or polar or nonpolar characteristics or the like using a method such as high performance liquid chromatography (HPLC; page 8, lines 21-24). The specification further teaches that a highly sensitive detection method such as mass spectrometry can be used to identify organ homing molecules (page 8, lines 11-15). Thus, the specification provides guidance to the skilled person, teaching that molecules can be separated, for example, based on molecular weight or polar or nonpolar characteristics and subsequently identified using a technique such as mass spectrometry.

The use of mass spectrometry to identify peptides is corroborated by the attached publication by Clauser et al., which states that mass spectrometry techniques "enable routine peptide sequencing from sample quantities of 100 fmol to 10 pmol," citing references 11 and 12 which were published in 1993 and 1994 (Exhibit 4 at page 5072, second column, first complete

paragraph). From the above, it is clear that in 1993 and 1994, in advance of the priority date of the present invention, one skilled in the art would routinely have been able to separate peptides by mass or charge and subsequently determine their peptide sequence using, for example, mass spectrometric analyses. In sum, Applicants submit that, in combination with knowledge of techniques which were routine in the art at the time the invention was made, the guidance in the specification would have been sufficient for one skilled in the art to practice the invention without undue experimentation.

In view of the above remarks, Applicants respectfully request that the Examiner reconsider and remove this ground for rejecting claims 1 to 3 and 6 to 7 as allegedly lacking enablement.

CONCLUSION

Applicants respectfully request that the Examiner enter the amendments and consider the remarks herein above. The Examiner is invited to call the undersigned agent or Cathryn Campbell if there are any questions relating to this application.

Respectfully submitted,

Date: May 26, 2004

Andrea L. Gashler
Registration No. 41,029
Telephone: (858) 535-9001
Facsimile: (858) 535-8949

McDERMOTT, WILL & EMERY
4370 La Jolla Village Drive
Suite 700
San Diego, California 92122

EXHIBIT 1

Combinatorial Drug Discovery: Which Methods Will Produce the Greatest Value?

David J. Ecker* and Stanley T. Crooke

Isis Pharmaceuticals, 2292 Faraday Avenue, Carlsbad, CA 92008. *Corresponding author: (e-mail: decker@isisph.com).

Combinatorial strategies are important new approaches to drug discovery, and it seems quite likely that they will result in the discovery of interesting potential pharmaceuticals. However, it is less clear whether combinatorial approaches will result in quantum advances in therapeutics. Nor is there general agreement about the factors most important in defining how combinatorial strategies will provide value to the discovery of lead and therapeutic compounds. In this review, we propose criteria that define the value of combinatorial strategies and categorize the various approaches by: (a) the type of chemical space to be searched, (b) the tactics employed to synthesize and screen libraries, and (c) the structures of individual molecules in libraries. We evaluate the strengths and weaknesses of the various strategies and suggest milestones that can help to track their success.

The two basic challenges of drug discovery are to identify a lead compound with desired activity and to optimize the lead compound to meet the criteria sufficient to proceed with development. Traditionally, identification of a lead compound has been achieved either by random screening of natural products or libraries of synthetic chemicals, or by modifications of chemicals with known physiological activities—usually the natural agonist for the targeted receptor. Advances in medicinal chemistry, structural and molecular biology, and molecular pharmacology have made the drug discovery process progressively more rational (i.e., rational drug design). The more that is known about the target for the drug, the more rationally the initial screening can be "biased," which usually leads to faster identification of the initial lead active. Once the lead compound is identified, medicinal chemistry can be used to optimize the desired activities, specifications, pharmacokinetic properties, and to attempt to reduce any toxicities.

Random screening of natural products is one of the technological bases of the modern pharmaceutical industry. Although it has resulted in many important drugs[1], the ratio of novel to previously discovered compounds has diminished with time. The development of novel synthetic libraries rekindled excitement about random screening as a paradigm for drug discovery. Several factors contribute to the enhanced interest. Perhaps most important are the new methods to rapidly synthesize and test libraries of novel chemicals. There is a growing perception that new chemical strategies are necessary to enable medicinal chemistry to take better and more rapid advantage of the numerous advances in the identification and understanding of molecular targets.

A broad variety of new synthesis and screening methods are currently grouped under the term "combinatorial." These methods include parallel chemical synthesis and testing of multiple individual compounds or compound mixtures in solution, synthesis, and testing of compounds on solid supports, and biochemical or organism-based synthesis of biological oligomers coupled to selection and amplification strategies. The various methods have proliferated rapidly; each with its putative advantages, disadvantages, and proponents.

Numerous development stage companies have focused on various combinatorial approaches and many major pharmaceutical companies have established combinatorial programs. These approaches have been discussed and compared in several reviews[2-9], and most scientists involved in drug discovery agree that combinatorial strategies are likely to result in the identification of increasing numbers of pharmacologically active lead molecules.

Far less consensus, however, has developed with regard to the answers to two critical and related questions: First, what are the factors that will determine whether combinatorial strategies create value as defined by the discovery of novel compounds with unique therapeutic properties? Second, which strategies and library synthesis and testing tactics are most likely to yield the greatest value?

Although answers to these two questions will emerge from the large array of experiments in progress, it is important to develop a conceptual framework to better define the process. The objectives of this article are to suggest a basis by which the potential value of combinatorial strategies can be assessed, and to identify milestones that will help provide an interim evaluation.

Criteria that Define the Value of Drug Discovery Strategies

The value of any drug discovery strategy can be determined by two criteria. First, does it result in an increase in the number of compounds that meet development criteria? Second, does it produce compounds with improved quality and performance in the clinic? The ultimate value of the different combinatorial drug discovery strategies can be judged by how well they perform in these two areas. Although the final evaluation can be made only after the products are delivered to the patients, there are some milestones that can be used to track the success of the various strategies (Table 1).

Success in increasing the number of drug candidates is the easier parameter to judge. This issue further reduces to two

TABLE 1. Criteria that define value.

Improvements in the quantity of drug candidates
∇ Rapid identification of potent leads
• Lead generated for targets with no structural information
• Inhibitors of multistep complex processes
∇ Rapid movement of leads into the clinic
• Leads that can move to development directly
• Rapid optimization
Improvements in the quality and performance of drug candidates
∇ Major improvements in receptor subtype specificity
∇ Activity against targets where classical approaches have failed
∇ Improvement in pharmacokinetics
∇ Low toxicity

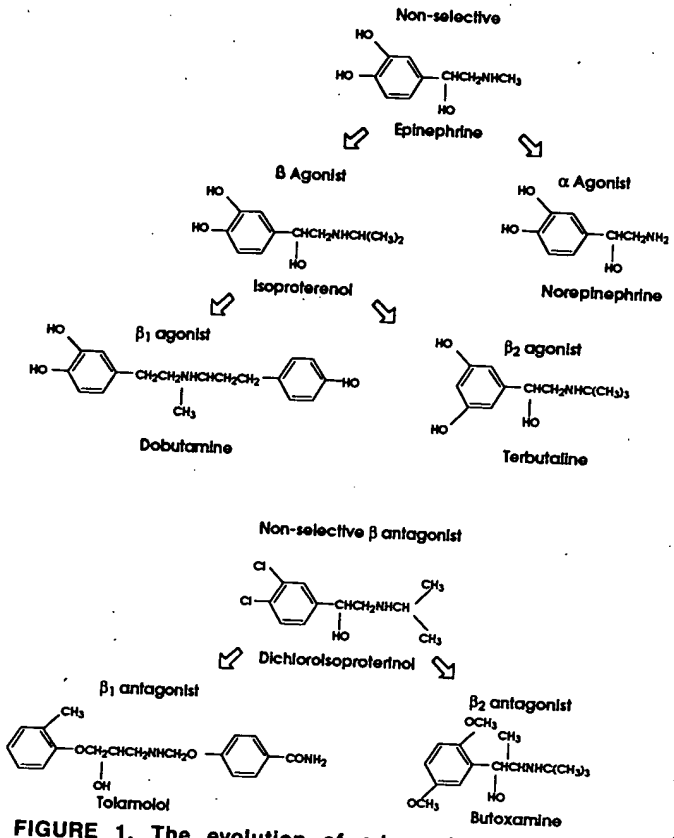


FIGURE 1. The evolution of adregenic receptor drug specificity.

questions: How rapidly are potent lead compounds identified, and how quickly can the lead compounds be channeled into clinical development? Success in the identification of potent leads can be evaluated early in the process of synthesis and screening. It may be particularly interesting to see how well various strategies perform in the identification of leads for targets where traditional methods have failed and where there is no structural information upon which to bias the search. Clearly value will be created by combinatorial methods if they succeed in delivering leads that have not been provided by other methods.

How easily and quickly the lead compounds can be moved into clinical development is the second half of the quantity question. Ideally, the compounds identified in the screenings should be directly acceptable as drugs, or only minor chemical modification should be required. This does not mean an optimization exercise should be eliminated. In fact, the more valuable discovery methods should facilitate rapid optimization of a lead. Methods that require a fundamental overhaul of the chemical composition of the lead will extend the time it takes for new products to reach the clinic. In addition, the original activity selected in screening is increasingly at risk as more extensive chemical modifications are introduced. In general, the methods that most rapidly produce lead compounds that can be moved quickly into clinical development will be the most valuable from a quantity perspective.

Improvements in the quality and performance of drug candidates in the clinic are harder to anticipate from early data. However, some fundamental factors can be considered. These include the role of the molecular targets for the drugs in the disease process and the specificity of a drug for its target. Other factors include the pharmacokinetic properties of the drug and the severity and incidence of the disease for which the drug is used. As with other approaches to drug discovery, combinatorial methods will create value if they provide a more efficient means of identifying selective pharmacological agents with acceptable pharmacokinetic properties designed to interact with important receptors.

What are the advantages of combinatorial chemistry that distinguish it from previous methods? Although combinatorial chemistry can be used to create novel structures, there is no reason to believe *a priori* that these structures will be any better than earlier randomly discovered structures. The specificity for the target is perhaps the single most important factor in determining whether a new compound will be a major therapeutic advance. Because combinatorial chemistry provides both a source of compounds to discover new leads and a more facile way to exhaustively search the structure space around the new lead, it can be argued that this approach offers more opportunities to find compounds with specific properties than other drug discovery strategies. Similar arguments can be applied to the other important parameters that make a useful drug such as the pharmacokinetic and toxicologic properties. However, it remains to be seen if these advantages will result in more than incremental improvements in the drug candidates. If combinatorial strategies simply increase the number of new leads discovered and optimized per unit time, then there is no reason *a priori* to believe that quantum advances in the therapeutic value of the new drugs discovered will be realized.

What is the Most Valuable Chemical Space to Search?

Given the variety in the types of structures that can be synthesized, a fundamental issue in combinatorial chemistry is: what kind of molecules should be made? Because of the experience of the pharmaceutical industry, we have a good idea what general kinds of compounds make useful drugs and combinatorial libraries based around these motifs will likely lead to the identification of new ones. This strategy is supported by the notion that combinatorial methods greatly increase the rate at which a structure space can be searched. If a certain structure has proven useful in the past, a more thorough exploration of similar structures should be a successful strategy.

However, the relative value of more exhaustively searching previously searched space versus searching novel space will be determined by the scope of whats out there to be discovered. An argument can be made that the search of structure space by the pharmaceutical industry represented a very small and highly biased sampling of a much more extensive universe of useful kinds of compounds. Historically, sampling of chemical space was based on the availability of compounds and synthetic capabilities of organic chemists. (It is interesting to speculate what the pharmaceutical industry would look like today if the earliest chemists of the industry were organometallic, bioinorganic, or polymer chemists.) Thus, it is important to consider arguments in favor of focusing combinatorial chemistries *away from* spaces previously searched by medicinal and natural product chemists.

Biological Specificity— The First Combinatorial Strategy

When confronted with the need to engineer specificity of action into chemicals to support complex biological processes, evolution generally chose to work with a relatively small number of monomers (e.g., 4 nucleosides, 20 amino acids) and created specificity by oligomerization. Relatively information-poor monomers, when oligomerized, result in unique shapes that are determined by the order of the monomers in the linear sequence. Additional specificity is sometimes obtained by the creation of branched or cyclic structures or by other post-oligomerization modifications. However, most of the information determining specificity is contained in the order of the

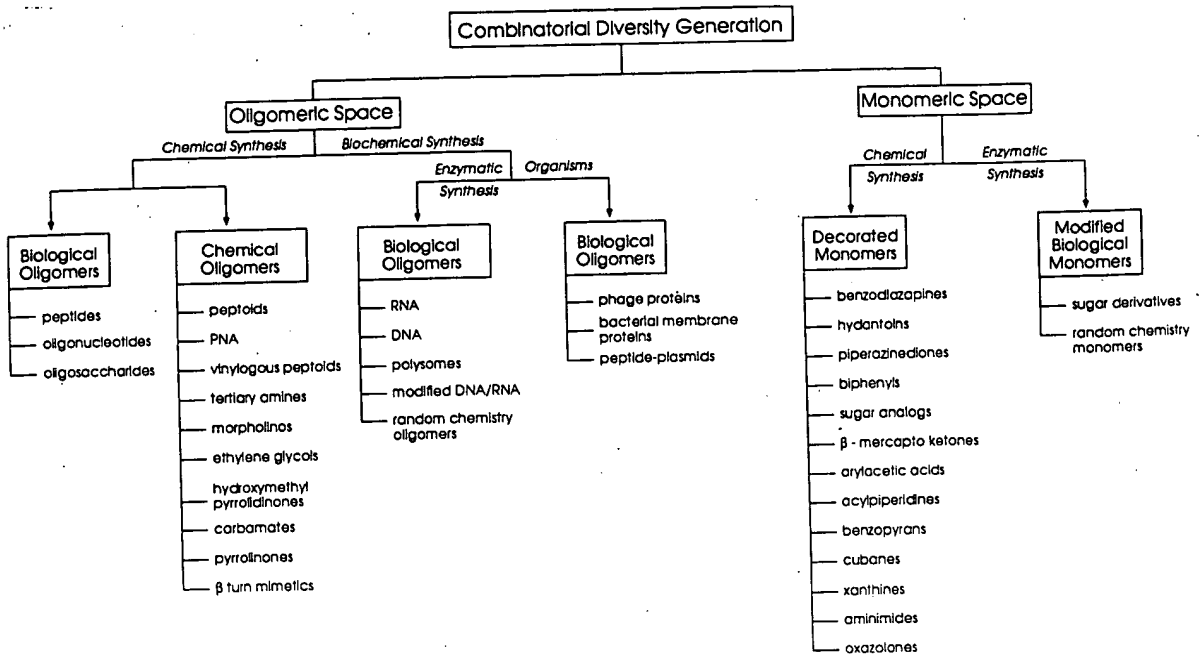


FIGURE 2. Combinatorial diversity generation.

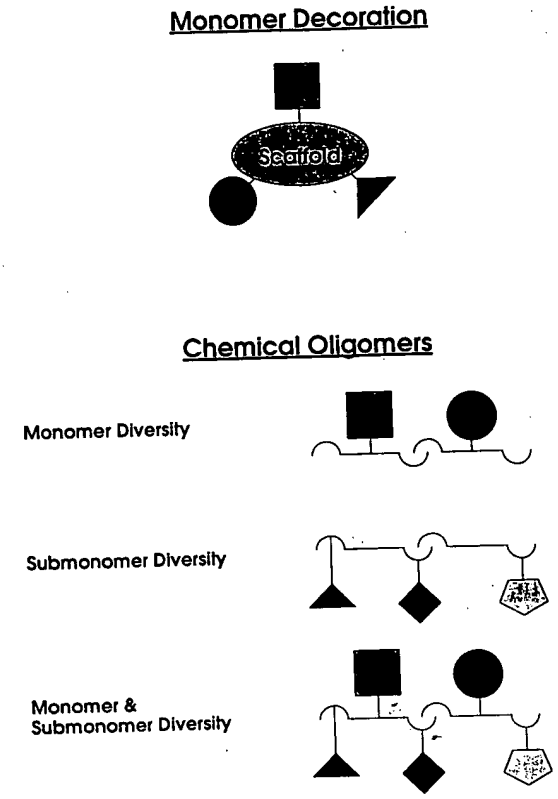


FIGURE 3. Combinatorial chemical motifs.

monomers, and post-oligomerization modifications provide the fine-tuning.

Specific biological interactions occur between molecules where at least one of the partners is an oligomer (e.g., epinephrine and its protein receptor). The most specific biological interactions usually involve two or more oligomers (e.g., transcription factor complexes). Information in biological oligomers is more than a text string-type of encoded data, such as the genetic sequence information stored in DNA. In structural RNAs and proteins, the specificity is engineered by the pattern of the display of the diverse side chains in space relative to each other and the common backbone to which they are attached.

Traditional Medicinal Chemistry—The Second Combinatorial Strategy

During the twentieth century, natural product and medicinal chemists have provided enormous libraries of compounds. Traditional drug discovery has been, in essence, a stepwise combinatorial search through the chemical structure space available to the practitioners of drug discovery. A common approach taken to optimize compounds to meet specific therapeutic objectives has been to add functionalities to lead structures, a process which we call "monomer decoration." For example, consider the history of adrenergic drug discovery (Fig. 1). Norepinephrine and epinephrine were shown to be natural agonists for the adrenergic receptors. As the properties of the adrenergic system and receptors were determined, it became clear that antagonists might have therapeutic value and then that isotype selective agonists and antagonists would be of considerable interest. Thus, during the past 40–50 years, the basic catecholamine structure has been progressively modified to create various classes of adrenergic agents. Extraordinary therapeutic advances and pharmacological tools enabled the progressive advances in understanding the adrenergic systems.

Although the term monomer decoration may sound simple, it represents a highly sophisticated set of advances in medicinal chemistry structure activity optimization including sidechain modification, isosteric replacement and structural constraining and stretching. It is clear that the strategies taken by medicinal chemists have differed dramatically from those taken by evolution. Novel combinatorial strategies can be categorized as those that facilitate the search of spaces traditionally searched by the medicinal chemists (i.e., approaches that facilitate monomer decoration) and those that facilitate the search of oligomeric space. These approaches can be categorized further on the basis of synthesis, compound identification and screening tactics.

Combinatorial Drug Discovery Motifs

A broad variety of very different strategies and tactics fall under the category of "combinatorial." Each method has certain strengths and weaknesses and some methods are differentially suited to specific purposes such as better understanding the origin of life, diagnostic testing kits, gene therapy and therapeu-

tic drugs. We will focus only on the issues relating to value in drug discovery.

Combinatorial drug discovery motifs are difficult to classify, but can be organized in a three-level hierarchy (Fig. 2): the type of space being searched, tactics to synthesize and screen, and the structures of the individual molecules in the libraries. The space searched can be divided into two groups: monomeric and oligomeric (Fig. 3). Although the two chemical groups do overlap, the chemical space for each is quite different.

Searching monomeric space. Decorated monomers are combinatorial extensions of classical medicinal chemistry. As in the development of adrenergic receptor drugs (Fig. 1), a basic scaffold with multiple sites for substitution is decorated with a variety of diverse functionalities. Placing the diverse functionalities in different positions relative to each other on the common scaffold creates new shapes.

The shape of the scaffold defines the general geometry of the space that can be searched. The molecular weight range of molecules in decorated monomer libraries will be determined by the choices of diverse functionalities that are attached. Generally, the molecular weight ranges are fairly tight, since all the molecules are based on a common scaffold.

Recently, a significant number of scaffold structures have been decorated in a combinatorial fashion (Fig. 2). The earliest versions produced relatively small arrays (~40) that were synthesized individually in parallel based on the benzodiazepine structure, which was already well known to have acceptable drug properties[7,10–12]. Also developed early were arrays of hydantoins and dipeptides[10]. As the technology improved, the numbers of compounds based on a single monomer scaffold increased to approximately 1000 and the compounds were synthesized in small mixtures. Recently a library of about 1000 piperazinediones was synthesized[13]. Also reported were small arrays of antioxidants[14], and a family of arylacetic acids[15]. Molecular scaffolds based on aminimide and oxazolone chemistries have been used to make arrays of individual compounds[16,17]. A series of decorated bis-benzamide phenols were described in which the scaffolds can be attached to each other[18,19]. Families of acylpiperidine and benzopyran compounds have been made on a support containing encoding binary tags[20]. Two different scaffolds based on cubane and xanthine containing four acid chloride functionalities were used to create a library of conformationally restricted structures[21,22].

All but the last two motifs were synthesized on a solid support. The range of organic reactions that can be successfully conducted on support-bound material is currently being explored[13,15,23]. The cubane and xanthine libraries were synthesized in solution and tested in complex mixtures. It is noteworthy that the decorated monomer motifs cited above have been screened using the whole range of screening formats including individual molecule testing, partial release from solid support, binary tags and iterative subset deconvolution. In all cases, compounds were tested in solution.

Enzymes can be used to modify biological monomers, either alone or in combination with chemical synthesis methods. A variety of modified simple sugars have been constructed using this strategy[24–26]. A strategy has been considered in which extremely complex mixtures of biological catalysts acting on themselves and on a basic set of monomers create supercritical reaction conditions that can generate billions of organic molecules[27].

Searching biological oligomeric space. Combinatorial drug leads can be obtained by using natural biological monomers and biological systems to assemble and sort them. In this approach, the principles of molecular evolution are applied to the discovery of drug lead structures[28]. Molecular diversity is

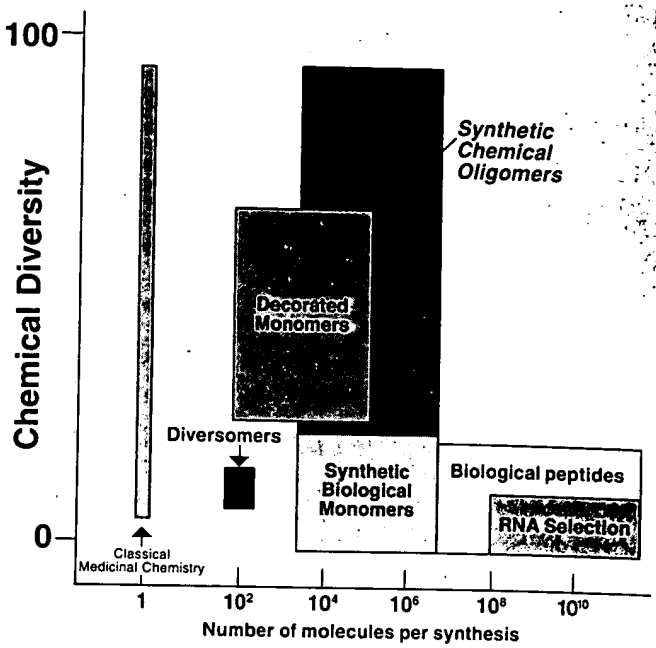


FIGURE 4. Diversity of various combinatorial libraries.

obtained by oligomerizing a relatively small number of monomers in a large number of ways. A selection strategy is imposed so that the most fit molecules survive a challenge that results in their amplification. In drug lead discovery, the challenge could be binding to a desired target. An amplification step is introduced so that molecules that meet the challenge are amplified and can be recovered. A measured number of mutations may be introduced in the hope of creating even more fit molecules and the process is cycled through multiple rounds.

Biological systems that have evolved to do this can be reconstituted *in vitro*. Currently this has been achieved with RNA[29–40], DNA[41–45] and proteins[46–52]. The wealth of success by many research groups in the identification of biological oligomers that bind to a variety of targets with high specificity is testimony to the power that *in vitro* biological systems have to generate a large number of shapes and the screening and sorting capacity of the selection scheme[29,33,34,38,42–44,53–60]. The ability to use directed molecular evolution to extensively search the structure space of biological oligomers is the great advantage of this method.

The difficult part comes after identification of the oligomer. Biological oligomers have limited value as drugs. They are not stable in biological systems, have poor pharmacokinetic properties and generally do not cross biological membranes, which coevolved with biological oligomers for the purpose of enclosing them. Since the whole purpose of drug therapy is to manage biology it is fundamentally difficult to override its homeostatic mechanisms with the same molecules that it already knows how to manage without making some chemical alterations.

Several strategies can be employed to overcome the natural limitations of biological oligomers after they have been selected. Mimicking biologically interesting peptide structures with pharmaceutically acceptable molecules has been a strategy pursued by many groups over the past 20 years[61–63]. RNA can be stabilized to a certain extent by chemical modification[64–70]. Generally, an RNA molecule with a desired activity is first identified and then modifications are systematically introduced to increase stability. However, any post-selection (back end) modification strategy runs the risk of losing the original activity as changes are made.

A front-end strategy is to introduce chemical modifications into the monomers that are tolerated by the synthesis and amplification system, and to perform the selection directly on the chemically modified oligomers. It has been shown, for example, that polymerases used in nucleic acid selection will accept certain 2′-sugar substitutions and pyrimidine substitutions in the 5-position[66,71]. This method has the advantage of preselecting only molecules with the desired activity. The limitation is that only a very small set of chemical substitutions are tolerated by polymerases.

Viruses, bacteria and mammalian cells can be used to generate tens of millions of peptides that can be screened for tight binding to any kind of receptor[47,51]. Mixtures of phage particles each containing unique peptide sequences can be prepared using a large number of random oligonucleotides and inserting them in the DNA that encodes various surface proteins[48,49,52,72-74]. Specific antibodies can be obtained from various selection strategies using phage mammalian cells or recombinant bacteria[46,75-81]. Combinations of selection and chemical modification methods further increase the level of diversity that can be generated using organisms[75,82]. Biologically generated structures can also be selected for catalytic function as well as for specific binding, and nucleic acids can be selected as well as peptides and proteins[81,83-88]. Strategies have been devised for selection in bacterial extracts in coupled transcription/translation systems by affinity selection of polysomes[89]. Selection of nucleic acids for antisense and triplex strategies has also been accomplished[54,85,90-92].

The likelihood that value will be produced from directed molecular evolution lies in the power of the search strategy to manage very large numbers of compounds. There is no comparably powerful search strategy available in any of the chemical methods. Undoubtedly, there will be value created by directed molecular evolution strategies for applications such as diagnostics, research reagents, and gene therapy. However, there are some high hurdles that must be cleared before large biological molecules can be modified into therapeutic drugs. It has yet to be demonstrated how long it will take for this strategy to produce clinical candidates of genuine therapeutic value.

Another approach for combinatorially sampling a large number of biological molecules is to chemically synthesize and screen them in solution by the use of analytical methods, iterative synthesis/screening strategies[93-99], or partial release from beads[100,101]. Alternatively, they can be screened on solid supports such as arrays[102-107], or multiple pins[108]. These methods have already demonstrated their value as tools for studying biological interactions and for generating structural leads for drug discovery. Chemical synthesis of biological molecules is powerful because chemical methods, rather than enzymes, are used to assemble the oligomers. This greatly facilitates the introduction of chemical modifications on the front end of the selection scheme and starts to blur the border between biological and chemical structures.

Searching chemical oligomeric space. While decorated monomer motifs follow the precedents of medicinal chemistry, chemical oligomers mimic the paradigm of biology: assemble a relatively small set of information-poor monomers in ways that create new shapes based upon the sequence order. The analogy between chemical and biological oligomers breaks down at the level of secondary and tertiary structure. It is generally considered important to keep the molecular weight of drug candidates low. Chemical oligomers of low molecular weight will not have the same capability as proteins for fold-back tertiary interactions over long distances. Therefore, other strategies are necessary for chemical oligomers to mimic discontinuous epitopes of proteins while maintaining low molecular weight. The desired degree of rigidity and structure will have to be engineered into the monomer units or introduced through branching or the creation of cyclic structures.

Decorated monomers frequently begin with motifs that already have attractive drug properties. With chemical oligomers, the task is to engineer the monomers so that the resulting oligomers have attractive pharmaceutical properties, i.e., stability in biological systems, desired physical properties, controlled structural rigidity, low molecular weight, and non-toxic component parts. Selecting active leads from a large repertoire of appropriately designed chemical oligomers, should reduce the steps to go from a lead compound to drug candidate in proportion to how well the initial monomers were designed.

The first generation of chemical oligomers were derived from variations on the techniques for chemical synthesis of biological oligomers. Analogs of peptides that use the amide linkage for oligomerization are the most frequently encountered chemical oligomers in combinatorial drug discovery. Variations include the use of D-amino acids[109,110], vinylogous versions[111], cyclic variations[112], and post-synthetically derivatized versions[113]. Peptides have been modified so that the variable functions or side chain is attached to the α amino group instead of the α carbon to produce N-substituted glycines or peptoids[114,115]. Likewise, PNA is another polyamide scaffold upon which diverse chemical functionalities can be displayed[116]. Unlike peptides, peptoids and PNA are not degraded by peptidases. Other motifs used in chemical oligomers include carbamates[117], pyrrolinones[118] and morpholinos[119,120]. Mimicking discontinuous epitopes of proteins with low molecular weight oligopeptides has been achieved recently by introducing a cyclic modification to produce a β-turn like structure[121].

Novel nucleotide analogs have been developed and oligomerized based upon the coupling reactions used in the chemical synthesis of DNA[122]. Newer generations of monomers which contain no sugars or nitrogenous bases but are coupled using phosphorous coupling chemistry have been developed. These include ethylene glycol or hydroxyprolinol linkers[122,123]. Each monomer unit is derivatized with a functionality that gives these oligomers unique physical and chemical properties.

A common feature of the first generation of chemical oligomers is a uniform repeating backbone. This is largely due to the fact that they are derived from variations of the molecules used in the chemical synthesis of biological oligomers. However, since chemical oligomers can be linked together by chemical reactions instead of enzymes, the practitioner of chemical oligomerization is not restricted to monomers that contain the same repeating linker unit. The ethylene glycol and hydroxyprolinol units are an example. The ethylene glycol linkers incorporate a relatively flexible alkyl chain while the hydroxyprolinol linkers are based on a more rigid pyrrolidine ring. These two types of linkage chemistries are compatible and, thus, can be used to create libraries of compounds that vary in rigidity. This capability is a new feature, and substantially increases the diversity that can be obtained.

Monomer and submonomer diversity. Two general strategies, monomers and submonomers, are available to introduce diverse functionalities into chemical oligomers (Fig. 3). Monomer diversity is achieved by the introduction of a diverse chemical functionality as part of the monomer unit. The monomers are then oligomerized together to create unique structures. Monomer diversity is the method of choice used by nature. The 20 amino acids and 4 nucleotides that compose proteins and RNA are examples of monomer-based diversity generation.

In the submonomer approach, the elongation of the chemical oligomer and the introduction of the diverse functional groups are accomplished in separate steps on the solid support. This has the advantage that separate monomers do not have to be preas-

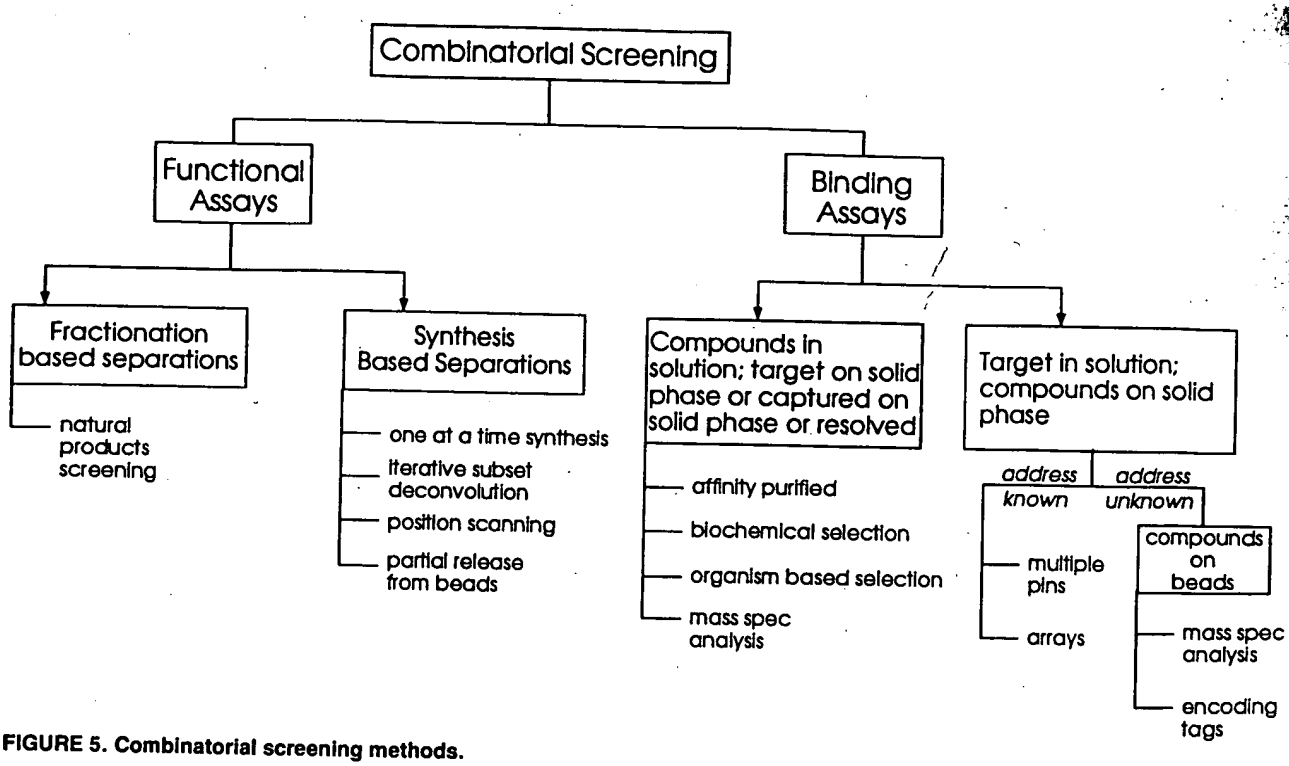


FIGURE 5. Combinatorial screening methods.

sembled, and allows the chemical manipulations to be automated. The synthesis of libraries of peptoids[114,115] illustrates this: the backbone portion of each unit is provided by coupling bromoacetic acid to the N-terminus of the oligomer, and the diverse functional groups are introduced by nucleophilic displacement of the bromide by a primary amine. The resulting N-substituted glycine terminus is now ready for the next elongation cycle. PNA is another chemical motif amenable to the generation of diversity through a submonomer approach. In PNA, one nitrogen serves as the point of attachment for the variable functionality while the other is used to link the components together. The latter position can serve as the point of attachment for additional functional groups.

We are currently applying the principles of both the monomer and the submonomer approaches for the preparation of phosphoramidate libraries (Hebert et al., in preparation). In this application, a hydrogen phosphonate monomer is coupled to the growing end of the oligomer, and the intermediate is converted to a phosphoramidate. A wide variety of amines can be employed in the reaction, and a large number of monomers have been prepared and are compatible with the synthesis conditions. By incorporating variability both in the linkage and in the backbone portion, a high density of functional groups can be achieved. Different backbone segments are expected to allow control over the presentation of the functional groups and create different shaped epitopes.

Overlap between monomeric and oligomeric space. Monomer decoration and chemical oligomerization are fundamentally very different approaches to searching space, but the shapes created by each approach can overlap. Non-peptidic monomeric scaffolds that attempt to mimic the shapes of peptides[62,63] are examples of overlap between the spaces. In a sense, much of the effort of rational drug design is an attempt to decorate monomeric scaffolds in ways that mimic (or complement) the shapes of biological oligomers. It will be interesting to see the results of experiments where unbiased libraries of chemical oligomers and decorated monomers are screened for activity against the same targets. Will the solutions provided by each approach be similar or dramatically different?

There is also opportunity for mixed mode searches. If two or more decorated monomers are linked together, the structure begins to cross over to a chemical oligomer motif with a non-repeating monomer unit. Likewise, small branched or cyclic oligomers may significantly overlap the same space as decorated monomers.

Useful diversity in combinatorial libraries. There are two distinct considerations in combinatorial library diversity: the number of molecules made per synthetic effort, and how different the molecules in the library are from each other (Fig. 4). Classical medicinal chemists produced one molecule for each synthetic effort. Although a single molecule cannot be considered "diverse," the molecule could fall anywhere on the landscape of all synthetically possible shapes. The first generation of decorated monomers, the diversomers[7], created more molecules at one time, but the molecules were very similar to one another. Newer generations of decorated monomers search shape space in clusters. Decorated monomers search the space defined by the geometry of the scaffold. The variety of ways to decorate any given scaffold provides for a very thorough search of the clump of space surrounding each scaffold.

Biological and chemical oligomers provide an even broader search of shape space than decorated monomers because the length of the oligomer can be varied to cover a greater variety of sizes within one library and there are more options for combinations. Synthetic chemical oligomers can have a higher diversity than synthetic biological oligomers for several reasons: more chemical monomers can be created, the "repeating unit" of chemical oligomers need not be homogeneous and synthetic submonomer diversity can be introduced in addition to monomer diversity. However, at this point the number of molecules that can be produced outstrips the capacity for screening.

Biological methods search the greatest total numbers due to their powers of selection and amplification. These very powerful search strategies allow a very thorough search of all possible shapes that can be created by biological oligomers. However, the number of monomers is restricted and the repeating units are homogeneous, so the overall chemical diversity can be considered to be lower than the chemical methods.

A distinct advantage of combinatorial chemistry compared with natural products is that the complexity as well as the composition of the mixture can be controlled. Choice of library complexity is an important factor in the design of libraries. A balance must be maintained between diversity and the concentration of each individual species. If libraries are too complex, the concentration of individual molecular species may be too low to measure a pharmacological response. Another consideration is that the split-synthesis method results in a unique compound synthesized on each bead of solid synthesis-support[124], so for all possible sequences to be represented, the number of beads must significantly outnumber the number of individual compounds produced[125]. Thus, although we have the ability to synthesize subsets containing millions of compounds, it is practical to limit complexity to the order of tens of thousands of compounds per library.

Screening Strategies for Chemical Libraries

Combinatorial screening strategies (Fig. 5) are nearly as diverse as the synthetic strategies. One of the major innovations of combinatorial technology is the strategic linking of screening and synthesis. In the past, natural product and chemical files were simply screened in functional assays: enzyme inhibition assays, receptor binding assays, whole cell assays, or animal assays. The pharmaceutical industry has developed a vast capacity to screen compounds or compound mixtures in functional assays. Moreover, functional assays select for the desired activity, while binding assays may select for compounds that bind the target protein, but fail to modulate its activity.

The simplest screening strategy currently considered "combinatorial" is the parallel chemical synthesis of many individual compounds, which subsequently are tested individually. Historically, this strategy has been used to screen chemical compound files. The innovation is completely on the synthesis end, and synthesis and screening are not strategically linked beyond the ability to conduct structure activity studies more rapidly than in the past. The major advantage is that this strategy is compatible with all existing assays. The major disadvantage is that the search rate is relatively limited. When screening capacity becomes limiting, individually synthesized compounds can always be pooled together and tested as a mixture. However, high-throughput screening methods are usually able to handle the yield of parallel chemical synthesis methods, in which case it is probably not worth invoking the added complications associated with screening mixtures, such as the need to deconvolute back to a single compound.

To begin to tap the true power of combinatorial methods in functional assays, compounds must be screened in mixtures. The related paradigm is natural product screening, which produced new compounds from a complex extract using analytical fractionation, retesting, purification and structural elucidation. Combinatorial methods have a number of conceptual advantages over natural product screening that can improve both the quality of the active compounds and the facility with which they are discovered. The chemical composition of active natural product is beyond control and requires structural elucidation to determine if it has any value. In many cases the structures cannot be synthesized or are too unstable to be useful drugs or cannot be analoged successfully. The composition of combinatorial libraries is under the complete control of the medicinal chemist and unstable or unsynthesizable compounds are excluded from the library from the beginning. While natural product extracts are not predisposed to revealing their structures, combinatorial libraries can be synthesized in a fashion that facilitates determining the structure of an active compound in a mixture without analytical separation, purification or structural characterization.

Strategic linking of synthesis and screening strategies in functional assays provides a means for determining the structures of active leads using various deconvolution strategies such as iterative subset deconvolution[93-99,126-130], position scanning[131,132], and partial release from solid support[100,101]. All of the above methods can be used in functional assays because compound mixtures can be tested in solution.

Screening mixtures of compound has a number of complications and caveats and it is far from clear which strategy is best. The success of screening extremely complex mixtures of compounds from natural product extracts in functional assays suggests that screening complex mixtures from combinatorial libraries should also be successful. However, the complexities produced by having more than one compound in the screen that arise in natural product drug discovery will also be encountered in screening combinatorial libraries. For example, there is always the chance for the activity of an active compound to be missed because of an interfering compound in the same mixture. There is the problem of multiple active compounds contributing to the measured activity in the mixture, which results in disappointment when after deconvolution no single molecule has high enough activity to be of value. Multiple actives may also make it difficult to find the best molecule in the mixture[133]. These complexities have always existed in natural product screening, and whether they will be more or less a factor in synthetic combinatorial libraries remains to be determined. An argument can be made that in combinatorial libraries these problems may be worse because the molecules that comprise the mixture are more similar to each other than they are in natural product extracts. Although this may be true for some combinatorial libraries, it is very difficult to make this comparison. Natural products often contain hundreds or thousands of different types of secondary metabolites such as alkaloids or steroid derivatives, yet specifically active representatives from each of these classes have been successfully detected and identified.

As the name implies, combinatorial drug discovery is a numbers game. The more drug discovery screens that compounds can be put through, the higher the probability of finding a lead. Thus, it is intuitively obvious that combinatorial strategies that are compatible with the many assays already established in the pharmaceutical industry will have a major advantage over strategies that require special assays to be developed for each target. Many man-years went into establishing, de-bugging and validating these assays by the industry, and compatibility with established assays is likely to be the most important issue relating to value from the standpoint of lead discovery. Lead optimization, on the other hand, is a process that should be more tolerant of an investment in assay development once it has been determined that there is an interesting lead.

Binding assays provide another useful screening tactic that has broad utility in drug discovery, diagnostic testing, DNA sequencing and basic biomedical research. Binding assays have special advantages such as the ability to control the relative concentrations of the target and compound libraries that avoid some of the pitfalls of functional inhibition assays. Binding assays can be divided into formats where the compounds are fixed on a solid phase or formats where the compounds start in solution and upon binding their target are captured on a solid phase or resolved through a chromatographic or mass-based separation technique. The solid support strategies usually have a built-in method to identify the most active compound such as addressable matricies[102-105,134,135] or pins[108]. Other strategies provide built-in handles to facilitate identification of the compound on a bead[101,124,136,137]. The recent dramatic improvements in mass spectrometry methods and micro-separation techniques have

greatly enhanced identification of compounds bound to a macromolecule even in the absence of tagging[138].

Conclusions

Combinatorial drug discovery methods have proliferated rapidly and there is a broad consensus that they will provide new leads for drug discovery. It is far less certain that combinatorial methods will provide compounds that represent major therapeutic advances. In this article, we have considered the issues in combinatorial drug discovery that are most important in creating value.

The value of any drug discovery strategy can be evaluated by determining the quantity and quality of the drugs that it produces. Improvements in the number of drug candidates is determined by both how rapidly lead compounds are discovered and how quickly they can be moved into the clinic. Improvements in the quality of drug candidates relates to activity against novel targets, increased specificity, improved pharmacokinetics and lower toxicity.

Whether combinatorial methods will be able to achieve these improvements will be determined by issues of chemistry, screening strategies and target applications (Table 2). The most important issues concerning chemistry are the novelty and quality of the space searched as well as the extent of the space searched. Search strategies in combinatorial drug discovery can be divided into monomeric and oligomeric strategies. Monomeric strategies follow the example of classical medicinal chemistry and search the same space that has been searched historically in a much more thorough fashion. The oligomeric strategy follows the paradigm of biology and searches novel space with a higher level of diversity. Biological oligomers can be searched in the greatest sheer numbers because of the power of directed molecular evolution as a selection strategy, but are severely disadvantaged by the amount of chemical modification required to make pharmaceutically acceptable compounds. Chemical oligomers have the highest overall diversity potential, but need to be proven to be pharmaceutically acceptable. The relative value of the monomeric and oligomeric strategies will be determined by the success of each in making discoveries.

Combinatorially discovered decorated monomers will likely yield the earliest drug candidates, but will result in a significant advance in therapeutics only if they target novel pathways or offer increases in specificity. The chemical oligomeric strategy is based on proven principles of biological specificity and provides the greatest opportunity to search away from previously searched space.

The most important issues in screening strategies are the breadth of assays amenable to the combinatorial method, the facility with which combinatorial libraries can be screened and the facility with which the method can identify a single compound from a large number of compounds. In order to tap the true power of combinatorial strategies, it would be desirable for large numbers of structures to be sampled and the synthetic and screening strategies must be strategically linked. With respect to these issues, methods that allow screening in the vast array of existing functional pharmaceutical assays coupled to synthesis-based separations, should they prove compatible, may have the edge over other methods. Binding-based separations have high utility when the target is known and available in high purity, and when there is interest in optimization of an interesting lead.

Individual companies involved in combinatorial drug discovery have limited resources and have to place their bets on the chemical strategies and screening methods that they judge will produce the greatest value. Ultimately, value will be determined in the laboratory and in the clinic as these methods are used to attempt to create new generations of drugs.

TABLE 2. Factors that determine value in combinatorial drug discovery methods.

Chemistry	Screening	Target Application
Novelty and quality of space searched • *Ability to continually make new structure classes* • *Ability to control physical and chemical properties of space searched*	Breadth of assays amenable to combinatorial method • *Compatibility with existing pharmaceutical screens*	Major therapeutic need
Extent of space searched • *Rate of synthesis of new structures* • *Ability to bias a search around a pre-selected space*	Facility with which libraries can be screened • *Minimum of hurdles or special methods*	Role of molecular target in the disease process
Time required to advance a lead structure to a drug candidate • *Minimal chemical modifications* • *Rapid optimization*	Facility of identification of a single compound from a large library	Compounds can achieve required specificity for target receptor
Cost of searching, optimizing leads, and scaling up drug candidate		

Acknowledgments

The authors would like to thank Oscar Acevedo, Frank Bennett, Tom Bruice, Dan Cook, Sue Freier, Rich Griffey, Normand Hebert, Debbie Robertson, Yogesh Sanghvi and Jackie Wyatt for their comments on this manuscript, Carol Collins and Laurie Capella for word processing, and David Pope for editorial assistance.

References

1. Spilker, B. 1989. Multinational Drug Companies: Issues in Drug Discovery and Development, Raven Press: New York.
2. Gordon, E. M., Barrett, R. W., Dower, W. J., Fodor, S. P. A. and Gallop, M. A. 1994. Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. J. Med. Chem. **37**:1385-1401.
3. Gallop, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. A. and Gordon, E. M. 1994. Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. J. Med. Chem. **37**:1233-1251.
4. Caporale, L. H. 1995. Chemical ecology: A view from the pharmaceutical industry. Proc. Natl. Acad. Sci. USA **92**:75-82.
5. Liskamp, R. M. J. 1994. Opportunities for new chemical libraries: Unnatural biopolymers and diversomers. ACIEE **33**:633-635.
6. Janda, K. D. 1994. Tagged versus untagged libraries: Methods for the generation and screening of combinatorial chemical libraries. Proc. Natl. Acad. Sci. USA **91**:10779-10785.
7. Bunin, B. A. and Ellman, J. A. 1992. A general and expedient method for the solid-phase synthesis of 1,4-benzodiazepine derivatives. J. Am. Chem. Soc. **114**:10997-10998.
8. Moos, W. H. and Green, G. D. 1993. Recent advances in the generation of molecular diversity. Annual Reports in Medicinal Chemistry **28**:315-324.
9. Ellington, A. D. 1994. Aptamers achieve the desired recognition. Current Biology **4**:427-429.
10. DeWitt, S. H., Kiely, J. S., Stankovic, C. J., Schroeder, M. C., Cody, D. M. R. and Pavia, M. R. 1993. "Diversomers": An approach to nonpeptide, nonoligomeric chemical diversity. Proc. Natl. Acad. Sci. USA **90**:6909-6913.
11. Bunin, B. A., Plunkett, M. J. and Ellman, J. A. 1994. The combinatorial synthesis and chemical and biological evaluation of a 1,4-benzodiazepine library. Proc. Natl. Acad. Sci. USA **91**:4708-4712.
12. DeWitt, S. H., Schroeder, M. C., Stankovic, C. J., Strode, J. E. and Czarnik, A. W. 1994. DIVERSOMER technology: Solid phase synthesis, automation, and integration for the generation of chemical diversity. Drug Development Research **33**:116-124.
13. Gordon, D. W. and Steele, J. 1995. Reductive alkylation on a solid phase: synthesis of a piperazinedione combinatorial library. Bioorganic & Medicinal Chemistry Letters **5**:47-50.
14. Kurth, M. J., Randall, L. A. A., Chen, C., Melander, C., Miller, R. B., McAlister, K., Reitz, G., Kang, R., Nakatsu, T. and Green, C. 1994. Library-based lead compound discovery: Antioxidants by an analogous synthesis/deconvolutive assay strategy. J. Org. Chem. **59**:5862-5864.
15. Backes, B. J. and Ellman, J. A. 1994. Carbon-carbon bond-forming methods

on solid support. Utilization of Kenner's "safety-catch" linker. J. Am. Chem. Soc. **116**:11171–11172.
16. Hogan, J. C., Jr. Inventor. 1993. Patent—Aminimide-containing molecules and materials as molecular recognition agents. USA. World Patent Application **9401102**.
17. Hogan, J. C., Jr. Inventor. 1994. Patent—Novel oxazolone derivatives are useful for synthesizing new compounds and fabricated materials which are able to recognize biological receptors enzymes, genetic materials, etc. USA. World Patent Application **9400509**.
18. Pavia, M. 1995. Abstract presented at the CHI conference, "Exploiting Molecular Diversity," January 23-25, San Diego, CA.
19. Sucholeiki, I. 1994. Solid-phase photochemical C-S bond cleavage of thioethers—A new approach to the solid-phase production of non-peptide molecules. Tetrahedron Lett. **35**:7307–7310.
20. Chabala, J. C. 1995. Abstract presented at the CHI conference, "Exploiting Molecular Diversity," January 23-25, San Diego, CA. abstract
21. Carell, T., Wintner, E. A. and Rebek, J., Jr. 1994. A solution-phase screening procedure for the isolation of active compounds from a library of molecules. Angew. Chem. Int. Ed. Engl. **33**:2061–2064.
22. Carell, T., Wintner, E. A., Bashir-Hashemi, A. and Rebek, Jr. 1994. A novel procedure for the synthesis of libraries containing small organic molecules. Angew. Chem. Int. Ed. Engl. **33**:2059–2061.
23. Chen, C., Ahlberg Randall, L. A., Miller, R. B., Jones, A. D. and Kurth, M. J. 1994. "Analogous" organic synthesis of small-compound libraries: Validation of combinatorial chemistry in small-molecule synthesis. J. Am. Chem. Soc. **116**:2661–2662.
24. Chou, W.-C., Chen, L., Fang, J.-M. and Wong, C.-H. 1994. A new route to deoxythiosugars based on aldolases. J. Am. Chem. Soc. **116**:6191–6194.
25. Henderson, I., Sharpless, K. B. and Wong, C.-H. 1994. Synthesis of carbohydrates via tandem use of the osmium-catalyzed asymmetric dihydroxylation and enzyme-catalyzed aldol addition reactions. J. Am. Chem. Soc. **116**:558–561.
26. Kajimoto, T., Chen, L., Liu, K. K.-C. and Wong, C.-H. 1991. Palladium-mediated stereocontrolled reductive amination of azido sugars prepared from enzymatic aldol condensation: A general approach to the synthesis of deoxy aza sugars. J. Am. Chem. Soc. **113**:6678–6680.
27. Kauffman, S. 1994. Protein mediated networks—Random chemistries. Ber. Bunsenges. Phys. Chem. **98**:1142–1147.
28. Joyce, G. F. 1992. Directed molecular evolution. Scientific American **267**:90–97.
29. Ellington, A. D. and Szostak, J. W. 1990. *In vitro* selection of RNA molecules that bind specific ligands. Nature **346**:818–822.
30. Szostak, J. W. 1992. *In vitro* genetics. TIBS **17**:89–93.
31. Tuerk, C. and Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science **249**:505–510.
32. Tuerk, C. and MacDougal-Waugh, S. 1993. *In vitro* evolution of functional nucleic acids: High-affinity RNA ligands of HIV-1 proteins. Gene **137**:33–39.
33. Giver, L., Bartel, D. P., Zapp, M. L., Green, M. R. and Ellington, A. D. 1993. Selection and design of high-affinity RNA ligands for HIV-1 Rev. Gene **137**:19–24.
34. Jenison, R. D., Gill, S. C., Pardi, A. and Polisky, B. 1994. High-resolution molecular discrimination by RNA. Science **263**:1425–1429.
35. Bartel, D. P. and Szostak, J. W. 1993. Isolation of new ribozymes from a large pool of random sequences. Science **261**:1411–1418.
36. Robertson, D. L. and Joyce, G. F. 1990. Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. Nature **344**:467–468.
37. Pan, T. and Uhlenbeck, O. C. 1992. *In vitro* selection of RNAs that undergo autolytic cleavage with Pb2+. Biochemistry **31**:3887–3895.
38. Tuerk, C., MacDougal, S. and Gold, L. 1992. RNA pseudoknots that inhibit human immunodeficiency virus Type 1 reverse transcriptase. Proc. Natl. Acad. Sci. **89**:6988–6992.
39. Joyce, G. F. 1989. Amplification, mutation and selection of catalytic RNA. Gene **82**:83–87.
40. Breaker, R. R. and Joyce, G. F. 1994. Combinatorial evolution of DNA molecules that cleave RNA. Chemistry and Biology **1**:223–299.
41. Beutel, B. A. and Gold, L. 1992. *In vitro* evolution of intrinsically bent DNA. J. Mol. Biol. **228**:803–812.
42. Ellington, A. D. and Szostak, J. W. 1992. Selection *in vitro* of single-stranded DNA molecules that fold into specific ligand-binding structures. Nature **355**:850–852.
43. Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathan, S. and Bolton, P. H. 1993. A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. Biochemistry **32**:1899–1904.
44. Griffin, L. C., Toole, J. J. and Leung, L. L. K. 1993. The discovery and characterization of a novel nucleotide-based thrombin inhibitor. Gene **137**:25–31.
45. Tsang, J. and Joyce, G. F. 1994. Evolutionary optimization of the catalytic properties of a DNA-cleaving ribozyme. Biochemistry **33**:5966–5973.
46. Osuna, J., Flores, H. and Soberon, X. 1994. Microbial systems and directed evolution of protein activities. Critical Reviews in Microbiology **20**:107–116.
47. Scott, J. K. and Craig, L. 1994. Random peptide libraries. Current Biology **5**:40–48.
48. Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F. and Gething, M.-J. H. 1993. Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. Cell **75**:717–728.
49. Dower, W. J. 1992. Phage power. Current Biology **2**:251–253.
50. Clackson, T. and Wells, J. A. 1994. *In vitro* selection from protein and peptide libraries. TIBTECH **173–184.**
51. Scott, J. K. and Smith, G. P. 1990. Searching for peptide ligands with an epitope library. Science **249**:386–390.
52. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. and Winter, G. 1991. Making antibody fragments using phage display libraries. Nature **352**:624–628.
53. Huizenga, D. E. and Szostak, J. W. 1995. A DNA Aptamer that binds adenosine and ADP. Biochemistry **34**:656–665.
54. Dehua, P., Ulrich, H. D. and Schultz, P. G. 1991. A combinatorial approach toward DNA recognition. Science **253**:1408–1411.
55. Mandelkow, E.-M. and Mandelkow, E. 1993. Tau as a marker for Alzheimer's disease. TIBS **18**:480–483.
56. Tsai, D. E., Kenan, D. J. and Keene, J. D. 1992. *In vitro* selection of an RNA epitope immunologically cross-reactive with a peptide. Proc. Natl. Acad. Sci. USA **89**:8864–8868.
57. Schneider, D., Tuerk, C. and Gold, L. 1992. Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. J. Mol. Biol. **228**:862–869.
58. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. and Toole, J. J. 1992. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. Nature **355**:564–566.
59. Chen, H. and Gold, L. 1994. Selection of high-affinity RNA ligands to reverse transcriptase: Inhibition of cDNA synthesis and RNase H activity. Biochemistry **33**:8746–8756.
60. Lorsch, J. R. and Szostak, J. W. 1994. *In vitro* selection of RNA aptamers specific for cyanocobalamin. Biochemistry **33**:973–982.
61. Kazmierski, W. M. 1994. Peptidomimetics and small-molecule drug design: towards improved bioavailability and *in vivo* stability. Trends in Biotechnology **12**:213–218.
62. Nicolaou, K. C., Salvino, J. M., Raynor, K., Pietranico, S., Reisine, T., Freidinger, R. M. and Hirschmann, R. 1989. Design and synthesis of a peptidomimetic employing β-D-glucose for scaffolding. Peptides **10**:881–884.
63. Hirschmann, R., Sprengeler, P. A., Kawasaki, T., Leahy, J. W., Shakespeare, W. C. and Smith, A. B., III. 1993. The versatile steroid nucleus: Design and synthesis of a peptidomimetic employing this novel scaffold. Tetrahedron **49**:3665–3676.
64. Usman, N., Beigelman, L., Draper, K., Gonzalez, C., Jensen, K., Karpeisky, A., Modak, A., Matulic-Adamic, J., DiRenzo, A., Haeberli, P., Tracz, D., Grimm, S., Wincott, F. and McSwiggen, J. 1994. Chemical modification of hammerhead ribozymes: activity and nuclease resistance. Nucleic Acids Symposium Series **31**:163–164.
65. Goodchild, J. 1992. Enhancement of ribozyme catalytic activity by a contiguous oligodeoxynucleotide (facilitator) and by 2'-O-methylation. Nucleic Acids Res. **20**:4607–4612.
66. Latham, J. A., Johnson, R. and Toole, J. J. 1994. The application of a modified nucleotide in aptamer selection: Novel thrombin aptamers containing 5-(1-pentynyl)-2'-deoxyuridine. Nucleic Acids Res. **22**:2817–2822.
67. Pieken, W. A., Olsen, D. B., Benseler, F., Aurup, H. and Eckstein, F. 1991. Kinetic characterization of ribonuclease-resistant 2'-modified hammerhead ribozymes. Science **253**:314–317.
68. Olsen, D. B., Benseler, F., Aurup, H., Pieken, W. A. and Eckstein, F. 1991. Study of a hammerhead ribozyme containing 2'-modified adenosine residues. Biochemistry **30**:9735–9741.
69. Heidenreich, O., Benseler, Fahrenholz, A. and Eckstein, F. 1994. High activity and stability of hammerhead ribozymes containing 2'-modified pyrimidine nucleosides and phosphorothioates. J. Biol. Chem. **269**:2131–2138.
70. Aurup, H., Tuschl, T., Benseler, F., Ludwig, J. and Eckstein, F. 1994. Oligonucleotide duplexes containing 2'-amino-2'-deoxycytidines: Thermal stability and chemical reactivity. Nucleic Acids Res. **22**:20–24.
71. Aurup, H., Williams, D. M. and Eckstein, F. 1992. 2'-Fluoro-and 2'-amino-2'-deoxynucleoside 5'-triphosphates as substrates for T7 RNA polymerase. Biochemistry **31**:9636–9641.
72. Lenstra, J. A., Erkens, J. H. F., Langeveld, J. G. A., Posthumus, W. P. A., Meloen, R. H., Gebauer, F., Correa, I., Enjuanes, L. and Stanley, K. K. 1992. Isolation of sequences from a random-sequence expression library that mimic viral epitopes. J. Immunol. Methods **152**:149–157.
73. Christian, R. B., Zuckermann, R. N., Kerr, J. M., Wang, L. and Malcolm, B. A. 1992. Simplified methods for construction, assessment and rapid screening of peptide libraries in bacteriophage. J. Mol. Biol. **227**:711–718.
74. Dedman, J. R., Kaetzel, M. A., Chang Chan, H., Nelson, D. J. and Jamieson, G. A.,Jr. 1993. Selection of targeted biological modifiers from a bacteriophage library of random peptides. The identification of novel calmodulin regulatory peptides. J. Biol. Chem. **268**:23025–23030.
75. Barbas, C. F., Bain, J. D., Hoekstra, D. M. and Lerner, R. A. 1992. Semisynthetic combinatorial antibody libraries: A chemical solution to the diversity problem. Proc. Natl. Acad. Sci. **89**:4457–4461.
76. Zebedee, S. L., Barbas, C. F., Hom, Y., Caothien, R. H., Graff, R., DeGraw, J., Pyati, J., LaPolla, R., Burton, D. R., Lerner, R. A. and Thornton, G. B. 1992. Human combinatorial antibody libraries to hepatitis B surface antigen. Proc. Natl. Acad. Sci. USA **89**:3175–3179.
77. Gram, H., Marconi, L., Barbas, C. F., Collet, T. A., Lerner, R. A. and Kang, A. S. 1992. *In vitro* selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. Pro. Natl. Acad. Sci. USA **89**:3576–3580.
78. Collet, T. A., Roben, P., O'Kennedy, R., Barbas, C. F., Burton, D. R. and Lerner, R. A. 1992. A binary plasmid system for shuffling combinatorial antibody libraries. Proc. Natl. Acad. Sci. USA **89**:10026–10030.
79. Munir, K. M., French, D. C. and Loeb, L. A. 1993. Thymidine kinase mutants obtained by random sequence selection. Proc. Natl. Acad. Sci. USA **90**:4012–4016.
80. Burton, D. R. 1993. Monoclonal antibodies from combinatorial libraries. Accounts Chem. Res. **26**:405–411.
81. Posner, B., Smiley, J., Lee, I. and Benkovic, S. 1994. Catalytic antibodies: perusing combinatorial libraries. TIBS **19**:145–150.
82. Barbas, C. F.,III, Amberg, W., Simoncsits, A., Jones, T. M. and Lerner, R. A. 1993. Selection of human anti-hapten antibodies from semisynthetic libraries. Gene **137**:57–62.

83. Eigen, M., Biebricher, C. K. and Gebinoga, M. 1991. The Hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. Biochemistry **30**:11005–11018.
84. Friedman, A. D., Triezenberg, S. J. and McKnight, S. L. 1988. Expression of a truncated viral trans-activator selectively impedes lytic infection by its cognate virus. Nature **335**:452–454.
85. Holzmayer, T. A., Pestov, D. G. and Roninson, I. B. 1992. Isolation of dominant negative mutants and inhibitory antisense RNA sequences by expression selection of random DNA fragments. Nucl. Acids Res. **20**:711–717.
86. Dube, D. K., Black, M. E., Munir, K. M. and Loeb, L. A. 1993. Selection of new biologically active molecules from random nucleotide sequences. Gene **137**:41–47.
87. Janda, K. D., Lo, C.-H. L., Li, T., Barbas, C. F.,III, Wirsching, P. and Lerner, R. A. 1994. Direct selection for a catalytic mechanism from combinatorial antibody libraries. Proc. Natl. Acad. Sci. USA **91**:2532–2536.
88. Brossalina, E. & Toulm, J.-J. 1993. A DNA hairpin as a target for antisense oligonucleotides. J. Am. Chem. Soc. **115**:796–797.
89. Mattheakis, L. C., Bhatt, R. R. and Dower, W. J. 1994. An *in vitro* polysome display system for identifying ligands from very large peptide libraries. Proc. Natl. Acad. Sci. USA **91**:9022–9026.
90. Noonberg, S. B., Scott, G. K., Garovoy, M. R., Benz, C. C. and Hunt, C. A. 1994. *In vivo* generation of highly abundant sequence-specific oligonucleotides for antisense and triplex gene regulation. Nucl. Acids Res. **22**:2830–2836.
91. Deiss, L. P. and Kimchi, A. 1991. A genetic tool used to identify thioredoxin as a mediator of a growth inhibitory signal. Science **252**:117–120.
92. Holzmayer, T. A., Pestov, D. G. and Roninson, I. B. 1992. Isolation of dominant negative mutants and inhibitory antisense RNA sequences by expression selection of random DNA fragments. Nucl. Acids Res. **20**:711–717.
93. Geysen, H. M. and Mason, T. J. 1993. Screening chemically synthesized peptide libraries for biologically-relevant molecules. Bioorganic & Medicinal Chemistry Letters **3**:397–404.
94. Erb, E., Janda, K. D. and Brenner, S. 1994. Recursive deconvolution of combinatorial chemical libraries. Proc. Natl. Acad. Sci. USA **91**:11422–11426.
95. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T. and Cuervo, J. H. 1991. Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. Nature **354**:84–86.
96. Eichler, J. and Houghten, R. A. 1993. Identification of substrate-analog trypsin inhibitors through the screening of synthetic peptide combinatorial libraries. Biochemistry **32**:11035–11041.
97. Houghten, R. A., Appel, J. R., Blondelle, S. E., Cuervo, J. H., Dooley, C. T. and Pinilla, C. 1992. The use of synthetic peptide combinatorial libraries for the identification of bioactive peptides. Peptide Research **5**:351–358.
98. Ecker, D. J., Vickers, T. A., Hanecak, R., Driver, V. and Anderson, K. 1993. Rational screening of oligonucleotide combinatorial libraries for drug discovery. Nucl. Acids Res. **21**:1853–1856.
99. Geysen, H. M., Rodda, S. J. and Mason, T. J. 1986. *A priori* delineation of a peptide which mimics a discontinuous antigenic determinant. Molecular Immunology **23**:709–715.
100. Salmon, S. E., Lam, K. S., Lebl, M., Kandola, A., Khattri, P. S., Wade, S., Pátek, M., Kocis, P., Krchnák, V., Thorpe, D. and Felder, S. 1993. Discovery of biologically active peptides in random libraries: Solution-phase testing after staged orthogonal release from resin beads. Proc. Natl. Acad. Sci. USA **90**:11708–11712.
101. Lebl, M., Pátek, M., Kocis, P., Krchnák, V., Hruby, V. J., Salmon, S. E. and Lam, K. S. 1993. Multiple release of equimolar amounts of peptides from a polymeric carrier using orthogonal linkage-cleavage chemistry. Int. J. Pept. Protein Res. **41**:201–203.
102. Fodor, S., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. and Solas, D. 1991. Light-directed, spatially addressable parallel chemical synthesis. Science **251**:767–773.
103. Maskos, U. and Southern, E. M. 1992. Parallel analysis of oligodeoxyribonucleotide (oligonucleotide) interactions. I. Analysis of factors influencing oligonucleotide duplex formation. Nucl. Acids Res. **20**:1675–1678.
104. Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P. and Fodor, S. P. A. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. USA **91**:5022–5026.
105. Maskos, U. and Southern, E. M. 1993. A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesised on a glass support. Nucleic Acids Res. **21**:4663–4669.
106. Maskos, U. and Southern, E. M. 1992. Oligonucleotide hybridisations on glass supports: A novel linker for oligonucleotide synthesis and hybridisation properties of oligonucleotides synthesised *in situ*. Nucl. Acids Res. **20**:1679–1684.
107. Matson, R. S., Rampal, J., Pentoney, S. L.,Jr., Anderson, P. D. and Coassin, P. 1995. Biopolymer synthesis on polypropylene supports: Oligonucleotide arrays. Anal. Biochem. **224**:110–116.
108. Maeji, N. J., Gordon, G., Bray, A. M. and Geysen, H. M. 1992. Simultaneous multiple synthesis of peptide-carrier conjugates. J. of Immunological Methods **146**:83–90.
109. Cooley, C. T., Chung, N. N., Wilkes, B. C., Schiller, P. W., Bidlack, J. M., Pasternak, G. W. and Houghten, R. A. 1994. An all D-amino acid opioid peptide with central analgesic activity from a combinatorial library. Science **266**:2019–2022.
110. Lam, K. S., Lebl, M., Krchnak, V., Wade, S., Abdul-Latif, F., Ferguson, R., Cuzzocrea, C. and Wertman, K. 1993. Discovery of D-amino-acid-containing ligands with selectide technology. Gene **137**:13–16.
111. Hagihara, M., Anthony, N. J., Stout, T. J., Clardy, J. and Schreiber, S. L. 1992. Vinylogous Polypeptides: An alternative peptide backbone. J. Am. Chem. Soc. **114**:6568–6570.
112. Eichler, J., Lucka, A. W. and Houghten, R. A. 1994. Cyclic peptide template combinatorial libraries: Synthesis and identification of chymotrypsin inhibitors. Peptide Res. **7**:300–307.
113. Ostresh, J. M., Husar, G. M., Blondelle, S. E., Dörner, B., Weber, P. A. and Houghten, R. A. 1994. "Libraries from libraries": Chemical transformation of combinatorial libraries to extend the range and repertoire of chemical diversity. Proc. Natl. Acad. Sci. USA **91**:11138–11142.
114. Simon, R. J., Kania, R. S., Zuckermann, R. N., Huebner, V. D., Jewell, D. A., Banville, S., Ng, S., Wang, L., Rosenberg, S., Marlowe, C. K., Spellmeyer, D. C., Tan, R., Frankel, A. D., Santi, D. V., Cohen, F. E. and Bartlett, P. A. 1992. Peptoids: A modular approach to drug discovery. Proc. Natl. Acad. Sci. USA **89**:9367–9371.
115. Zuckermann, R. N., Martin, E. J., Spellmeyer, D. C., Stauber, G. B., Shoemaker, K. R., Kerr, J. M., Figliozzi, G. M., Goff, D. A., Siani, M. A., Simon, R. J., Banville, S. C., Brown, E. G., Wang, L., Richter, L. S. and Moos, W. H. 1994. Discovery of nanomolar ligands for 7-transmembrane G-protein-coupled receptors from a diverse N-(substituted)glycine peptoid library. J. Med. Chem. **37**:2678–2685.
116. Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B. and Nielsen, P. E. 1993. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. Nature **365**:566–568.
117. Cho, C. Y., Moran, E. J., Cherry, S. R., Stephans, J. C., Fodor, S. P. A., Adams, C. L., Sundaram, A., Jacobs, J. W. and Schultz, P. G. 1993. An unnatural biopolymer. Science **261**:1303–1305.
118. Smith, A. B., III, Keenan, T. P., Holcomb, R. C., Sprengler, P. A., Guzman, M. C., Wood, J. L., Carroll, P. J. and Hirschmann, R. 1992. Design, synthesis and crystal structure of a pyrrolinone-based peptidomimetic possessing the conformation of a β-strand: Potential applications to the design of novel inhibitors of proteolytic enzymes. J. Am. Chem. Soc. **114**:10672–10674.
119. Stirchak, E. P., Summerton, J. E. and Weller, D. D. 1987. Uncharged stereoregular nucleic acid analogs. Synthesis of a cytosine-containing oligomer with carbamate internucleotide linkages. J. Org. Chem. **52**:4202–4206.
120. Stichak, E. P., Summerton, J. E. and Weller, D. D. 1989. Uncharged stereoregular nucleic acid analogs: 2. Morpholino nucleoside oligomers with carbamate internucleotide linkages. Nucl. Acids Res. **17**:6129–6141.
121. Virgilio, A. A. and Ellman, J. A. 1994. Synthesis of β-turn mimetics II. J. Am. Chem. Soc. **114**:11580–11581.
122. Davis, P. W., Vickers, T. A., Wyatt, J. R., Lingardo, L., Acevedo, O., DeBaets, E., Guinosso, C., Sanghvi, Y., Cook, P. D. and Ecker, D. E. 1995. Novel inhibitors of LTB4 and PLA2 discovered from novel combinatorial oligomer libraries. Submitted to J. Am. Chem. Soc.
123. Hebert, N., Davis, P. W., DeBaets, E. L. and Acevedo, O. L. 1994. Synthesis of N-substituted hydroxyprolinol phosphoramidites for the preparation of combinatorial libraries. Tetrahedron Lett. **35**:9509–9512.
124. Salmon, S. E., Lam, K. S., Felder, S., Yeoman, H., Schlessinger, J., Ullrich, A., Krchnak, V. and Lebl, M. 1994. One bead, one chemical compound: Use of the selectide process for anticancer drug discovery. Acta Oncologica **33**:127–131.
125. Burgess, K., Liaw, A. I. and Wang, N. 1994. Combinatorial technologies involving reiterative division/coupling/recombination: Statistical considerations. J. Med. Chem. **37**:2985–2987.
126. Blake, J. and Litzi-Davis, L. 1992. Evaluation of peptide libraries: An iterative strategy to analyze the reactivity of peptide mixtures with antibodies. Bioconjugate Chem. **3**:510–513.
127. Dooley, C. T., Hope, S. and Houghten, R. A. 1994. Rapid identification of novel opioid peptides from an N-acetylated synthetic combinatorial library. Regul. Pept. **54**:87–88.
128. Pinilla, C., Appel, J. R. and Houghten, R. A. 1993. Functional importance of amino acid residues making up peptide antigenic determinants. Mol. Immunol. **30**:577–585.
129. Wyatt, J. R., Vickers, T. A., Roberson, J. L., Buckheit, R. W., Jr., Klimkait, T., DeBaets, E., Davis, P. W., Rayner, B., Imbach, J. L. and Ecker, D. J. 1994. Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. Proc. Natl. Acad. Sci. USA **91**:1356–1360.
130. Owens, R. A., Gesellchen, P. D., Houchins, B. J. and DiMarchi, R. D. 1991. The rapid identification of HIV protease inhibitors through the synthesis and screening of defined peptide mixtures. Biochem. Biophys. Res. Commun. **181**:402–408.
131. Dooley, C. T. and Houghten, R. A. 1993. The use of positional scanning synthetic peptide combinatorial libraries for the rapid determination of opioid receptor ligands. Life Sci. **52**:1509–1517.
132. Pinilla, C., Appel, J. R., Blanc, P. and Houghten, R. A. 1992. Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries. BioTechniques **13**:901–905.
133. Freier, S. M., Konings, D. A. M., Wyatt, J. R. and Ecker, D. J. 1995. Deconvolution of combinatorial libraries for drug discovery: A model system. Journal of Medicinal Chemistry **38**:344–352.
134. Drmanac, R., Drmanac, S., Strezoska, Z., Paunesku, T., Labat, I., Zeremski, M., Snoddy, J., Funkhouser, W. K., Koop, B., Hood, L. and Crkvenjakov, R. 1993. DNA sequence determination by hybridization: A strategy for efficient large-scale sequencing. Science **260**:1649–1652.
135. Southern, E. M., Case-Green, S. C., Elder, J. K., Johnson, M., Mir, K. U., Wang, L. and Williams, J. C. 1994. Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. Nucl. Acid Res. **22**:1368–1373.
136. Nestler, H. P., Bartlett, P. A. and Still, W. C. 1994. A general method for molecular tagging of encoded combinatorial chemistry libraries. J. Org. Chem. **59**:4723–4724.
137. Borchardt, A. and Still, W. C. 1994. Synthetic receptor binding elucidated with an encoded combinatorial library. J. Am. Chem. Soc. **116**:373–374.
138. Chu, Y.-H., Avila, L. Z., Biebuyck, H. A. and Whitesides, G. M. 1993. Using affinity capillary electrophoresis to identify the peptide in a peptide library that binds most tightly to vancomycin. J. Org. Chem. **58**:648–652.

EXHIBIT 2

JOURNAL OF
MEDICINAL
CHEMISTRY

Volume 37, Number 10 May 13, 1994

Perspective

Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies, and Future Directions[1]

Eric M. Gordon,* Ronald W. Barrett, William J. Dower, Stephen P. A. Fodor, and Mark A. Gallop

Affymax Research Institute, 4001 Miranda Avenue, Palo Alto, California 94304

Received March 1, 1994

Biographies

Ronald W. Barrett received his Ph.D. in Pharmacology from Rutgers University in 1983 and did postdoctoral work at the Addiction Research Foundation in Palo Alto, CA, in the field of opiate receptors. From 1986 to 1989, he worked in the Neuroscience Research Area at Abbott Laboratories in Abbott Park, IL. He joined Affymax in 1989 where he established the Receptor Pharmacology Group. Dr. Barrett is presently Vice President and Director of Receptor Pharmacology.

William J. Dower received his Ph.D. in Biology from the University of California, San Diego, and did postdoctoral work at Stanford University studying steroid control of specific mRNA metabolism and the amplification of genes mediated by chemotherapeutic agents. He joined Bio-Rad Laboratories in 1984 where as a member of the newly formed molecular biology unit, he introduced the electroporation technique for the high-efficiency transformation of bacteria. He joined Affymax in 1989 to establish a molecular biology research group. Dr. Dower is Director of Molecular Biology at Affymax.

Stephen P. A. Fodor received his Ph.D. in Chemistry at Princeton University and was an NIH postdoctoral fellow in Chemistry at the University of California, Berkeley. He joined Affymax in 1989 as a Staff Scientist in Optical Technologies, becoming Director of Physical Sciences. Dr. Fodor's group led the development of new technologies, merging photolithography with combinatorial solid-phase chemistry. In 1993 he joined Affymetrix as Scientific Director, where he is currently using oligonucleotide arrays to study a variety of DNA molecular recognition processes.

Mark A. Gallop received his Ph.D. from the University of Cambridge, England, for research in the area of organo transition metal chemistry and did postdoctoral work in the laboratories of Peter G. Schultz at the University of California, Berkeley. He joined Affymax in 1990 and is currently a Senior Scientist and leader of the Combinatorial Chemistry Group.

Eric M. Gordon received his Ph.D. in 1973 from the University of Wisconsin—Madison and did postdoctoral work at Yale University. His research interests include the rational design of enzyme inhibitors, the chemistry of amino acids, peptides, and natural products, and combinatorial chemistry. Dr. Gordon has authored 150 papers and U.S. patents in these areas. From 1974 to 1992 he worked at the Squibb Institute for Medical Research (subsequently the Bristol-Myers Squibb Pharmaceutical Research Institute) in Princeton, NJ, most recently as a Director of Medicinal Chemistry. In 1992, he joined the Affymax Research Institute where he is currently Vice President of Research and Director of Chemistry.

A. Combinatorial Organic Synthesis

The notion of creating huge, searchable libraries of small organic molecules is unprecedented in medicinal chemistry, and the possibility of doing so has recently captured the imagination of the drug-discovery community. The conventional paradigm of small molecule lead development, in which a compound undergoes many rounds of individualized, hand-crafted modification and biological testing en route to drug candidacy, will likely be dramatically accelerated by the application of combinatorial chemistry technologies to mass-produce and evaluate lead analogs. The ability to harness molecular diversity techniques as tools for lead discovery offers an unparalleled opportunity for medicinal chemistry to expand the breadth and scope of molecular structures that may be screened for biological activity. Widespread availability of collections of highly diverse small-molecule libraries should provide an opportunity to assess the impact of combinatorial organic synthesis on new-lead discovery. In this section of part 2, some of the issues which confront the practitioner of combinatorial organic synthesis, as they relate to the problems of molecular recognition in general and medicinal chemistry in particular, will be analyzed.

Issues in Practicing Combinatorial Organic Synthesis

Combinatorial organic synthesis (COS) presents somewhat of an intellectual inversion of the past 50 years of synthetic organic chemistry. The chemist of the Woodwardian era was interested in a masterful and carefully

0022-2623/94/1837-1385$04.50/0

plotted natural product synthesis of a complex entity of known structure. Reactions were more often employed or developed to solve specific challenges rather than to provide generic methodologies. Rigorous control of reaction pathways, stereochemistry, and regiochemistry, and the exclusion of all but the desired diastereomer were obligatory in a faithful rendering of the technique. In contradistinction to natural product total synthesis, rather than generating a single, specific entity, the goals of COS are to create *populations* of molecular structures. Rather than exercising complete control, the combinatorial chemist, while maintaining high reaction efficiency and relative reactive compatibility, may actually seek to create situations and apply strategies in which stereochemical/regiochemical control is relaxed. This must be achieved while remaining cognizant of the impact these factors may have on the stoichiometry of the resulting libary and its design and ultimate use. Hence, the combinatorial chemist seeks to apply a series of Woodwardian reactions (reliable, high yielding) that operate generically on a diverse set of building blocks to provide a multitude of related products.

Criteria for Library Design. The primary objectives of producing small-molecule libraries by COS are to provide collections of compounds suitable for both drug-discovery screening and drug-development optimization. When complete, the combinatorial drug-discovery exercise should have created a stable population of low molecular weight entities, free of reactive and toxicity-causing functionality. While a paramount medicinal chemistry design criteria for small-molecule-libary construction might be that the *products* of diversity generation (individual library members) should "look" like drug leads, of still greater importance is that the library actually contains compounds capable of interacting at some detectable level with the biological target of interest. When small-molecule leads for a target have been previously defined (e.g., benzodiazepine ligands for a peptide or other G-protein-coupled receptor, transition-state inhibitors for a protease), the notion of searching for more potent derivatives among libraries combinatorially enriched in specific pharmacophore analogs is an obvious tactic to pursue. However, as the universe of well-defined macromolecular drug targets continues to expand through the impact of molecular cloning, the problem of identifying new pharmacophores capable of modulating the various interactions of peptides, proteins, carbohydrates, oligonucleotides, or lipids at these sites will also be intensified.

Will "rules" about the types of libraries that may prove generally useful in ligand discovery be discovered? Although the field of molecular diversity has not yet matured to the point where substantial insight into this question is forthcoming, it is intuitively obvious that small-molecule libraries, whose members structurally resemble historical leads, should provide a fertile reservoir of potential molecular diversity. Tangential to this, natural products aside, numerous historical drug leads were derived simply because synthetic routes to these molecules were readily available. It is likely that early-stage COS will be limited by applicable chemistry and that this will necessarily focus work toward traditional leads, whose syntheses are known and well-documented.

The successful identification from recombinant libraries of L-amino acid-based peptide ligands that inhibit protein–peptide, protein–protein, and protein–carbohydrate interactions suggests there is broad utility in screening large libraries of peptidic compounds. It remains to be seen whether collections of other random molecular structures that are quantitatively as diverse as existing peptide libraries prove in *de novo* ligand discovery to include the "pharmacophores of the future".

Ligand rigidity may be another important parameter to consider in the course of library design. The incorporation of conformational constraints into flexible lead molecules has emerged as a powerful strategy to enhance ligand potency and/or selectivity, particularly in the field of peptidomimetic medicinal chemistry.[2-10] Nevertheless, with regard to library design, conformational restriction may act as a two-edged sword: an inappropriate constraint is likely to abrogate the modest but perhaps detectable activity of a more flexible analog, which could, in a secondary library, be systematically constrained. From the point of view of random screening, it remains to be determined whether useful leads will arise more frequently from libraries of rigidified or flexible structures. Data from the evaluation of cyclic peptide libraries in both synthetic and recombinant systems may provide some important insights into this issue. A number of methods have been recently described for on-resin cyclization of peptides through both main-chain and side-chain functional groups.[6-10,68] At present, a portfolio of libraries containing both conformationally rigid and relaxed molecular diversity seems most appropriate. A longer range solution might be to moderate the high risk of conformational restriction by creating very large populations of semirigid molecular arrays, comprising structural families that collectively sample as completely as possible all regions of conformational space.

Characterization. The usual measures of evaluating success in organic synthesis may lose meaning in COS. The classical notions of such fundamental concepts as purity/homogeneity, yield, exact product structure, relative and absolute stereochemical control, specific physical properties are less relevant when applied to a broad population of molecules (of course they may become quite relevant as individuals emerge from a selection process). Additionally, the analytical mainstays of the synthetic organic chemist, such as NMR and IR, may become obviated. The NMR spectrum of a 10 000-component library mixture is not diagnostic. The loss of these powerful tools requires that compensating technologies be developed. A major dilemma of COS is the difficulty of confirming the degree to which the expected chemistry has proceeded on the entire population of substrate molecules. Several groups have recently reported on the use of electrospray mass spectrometry as a technique for evaluating the bulk composition of diverse peptide libaries.[11,12] Gross synthetic discrepancies, such as incomplete protecting group removal, may be detectable by mass analysis, providing an opportunity to optimize the library synthesis protocols. In the characterization of combinatorial products, the presence of "byproducts" (in COS, unexpected products), combined with the difficulty of detecting these compounds, will cause problems if one mistakenly concludes that a screening hit is the expected product. This section will conclude by offering a potential solution to this problem.

Efficiency/Automation. Among the chemical criteria relevant for small-molecule-library design is the efficiency of diversity creation. The assembly of most small molecules reduces to the intercombination of only three to

five building blocks of molecular weight ~150 each. Synthetic reactions capable of combining numerous building blocks simultaneously constitute a highly efficient form of diversity generation. Thus the Ugi four-component reaction has a high combinatorial efficiency since building blocks of four families (amines, carbonyl compounds, isocyanides, and suitable acid components) are linked simultaneously to afford α-amino acid derivatives. In contrast, peptide chemistry traditionally links two building blocks at a time. In both the broad screening and the lead analoging modes, a longer range question pertains to the ability of the chemistry to eventually be automated. Once the key decisions and overall strategy have been determined, much of the actual chemistry is repetitive in nature. Machines will continue to be constructed to capitalize on this and libraries will be assembled under computer control.[13-15]

Quantity and Quality of Diversity. While the "quantity of diversity" that is experimentally accessible can be dictated by the number of building blocks in the basis set and by the number of synthetic operations applied, or able to be applied (see part 1[1]), the practical limitations on library size are most generally imposed by the format within which the diversity is created and evaluated. A small number of building blocks subjected to many synthetic steps will yield high (numerical) diversity; however the products of these reactions may be relatively large molecules, not well-suited for lead development as traditionally administered therapeutics. Thus, as the combinatorial process proceeds, an opportunity window may exist in which the bulk of the library possesses properties which standard medicinal chemistry usually seeks in small-molecule drug discovery (MW < 700, solubility, etc.). Continued application of the combinatorial process will lead to product libraries containing larger molecules (composed of more building blocks) wherein the individual library members have "outgrown" the classical criteria of a lead-drug molecule.

In surveying the historical landscape of drug discovery, there are particular pharmacophores or structural arrays which periodically surface far in excess of random chance (benzodiazepines, β-lactams, imidazoles, phenethylamines, etc.).† A review of recent successes in the era of "rational drug design" suggests that certain molecular concatenations—protein turn mimetics, conformationally restricted amino acids, transition-state analogs, dipeptide isosteres, molecular scaffolds, designed elements for enzyme inhibition—are often found in the medicinal chemistry of lead compound development. In consideration of the molecular structures which have left their mark on modern medicinal chemistry, one might conclude that the drug-discovery process is impacted not only by the sheer quantity of diversity surveyed, but additionally by the more subjective "quality" of diversity that is evaluated. Different organizations and individuals will certainly bring a wide variety of criteria to the subjective appraisal process, depending on style, experience, and bias. It may be speculated that the quality of diversity will be influenced by the sophistication of the building blocks originally introduced into the combining system (library bias on the part of the medicinal chemist) and the extent to which molecular substructures of the building blocks can be assembled in diverse, spatial (3-D) relationships. Thus the collected expertise of medicinal chemical knowledge may be used to "hyperevolve" or "bias" the libary by the planned introduction of commonly evolved elements; these elements are "retrocombinatorial synthons" of many known bioactive classes. Thus the building block basis set must be judiciously chosen and carefully attuned to the collected knowledge historically amassed in drug discovery.

† An interesting aside regarding these important substructures is that development of "generic" syntheses of key pharmacophores ultimately enabled facile generation of many analogs. Concurrently or subsequently, diverse biological activities were found among these compound classes. In a sense, this is suggestive of combinatorial chemistry, except the crucial molecules were made serially rather than in a parallel/combinatorial high throughput fashion.

Issues in the Selection of Building-Block Sets

The acquisition of a building-block library can be a major time and resource investment, and the eventual decision of which type of chemical building blocks to utilize places limits on the universe of structural diversity which ultimately can be explored. Depending on the specific objective, important building-block criteria include the availability of a large number of diverse, fairly complex, easily accessible starting materials. These may be either commercially available or prepared in a few steps from commercial materials. Members of a building block set should reflect a broad array of physicochemical properties, functionality, charge, conformation, etc. Building blocks may be chiral, achiral, or racemic. Certain building-block families have what may be termed a high "combinatorial potential". This relates to the high density per carbon atom of reactive functionality which can participate in new covalent combinations. For example, monosaccharides have high combinatorial potential since the high density of available hydroxyl groups leads to many potential connecting permutations. In addition to polymer formation, the high combinatorial potential of such types of building blocks may also be exploited as scaffolds for the generation of diversity (*vide infra*).

Synthetic Strategy

An important strategic element in combinatorial library synthesis is the degree of reliability of the ligand synthesis chemistry. What is the likelihood of general synthetic success with a particular reaction? The nature of combinatorial reactions, which must proceed in the face of a broad range of functionality on a multitude of substrates and where the products are difficult to analyze individually, demands that, in selection of synthetic methodologies, greater weight must be given to reaction sequences with reliable, predictable outcomes. A more subtle question revolves around the number of synthetic options available in the course of diversity generation. For example, a synthetic strategy structured in such a way that, as the process proceeded, new combinatorial possibilities opened up, would be preferable to having options narrow, especially if the goal was generating a maximum of structural diversity.

As previously noted, there are two distinct themes that must be considered for the successful application of combinatorial technologies to ligand discovery and optimization, *viz.* broad-based screening and directed chemical analoging. The issues underlying conceptual design, as well as the synthetic strategies utilized in construction of these different classes of libraries, are noteworthy and are summarized in Figure 1. Building block requirements for undertaking broad and narrow diversity searches differ markedly. The search for an initial lead molecule may be essentially a random screening exercise, where the em-

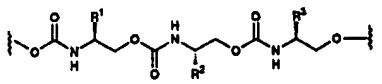
Broad Screening	Chemical Analoging/Optimization
huge size library	modest size library
broadest structural diversity	relatively narrow structural diversity
no special initial structure goal	specific structural goal
any building blocks	specific retrocombinatorial building blocks
undefined order of reaction	specific order of combination
flexible synthetic strategy	well defined synthetic strategy
site of tether not crucial	tether crucial-build in redundancy
ligand possibly uncouplable	ligand should be releasable
single selection evolution	cumulative selection evolution

Figure 1. Combinatorial chemistry: comparison of two major themes.

phasis is on exposing a macromolecular drug target to the maximum possible structural diversity. The objective is to identify a ligand of significant affinity for the target, the exact ligand structure and its detailed characteristics at this point are not relevant: in fact any molecule will do. An approach to generating highly diverse libraries for use in medicinal chemistry might favor using building blocks which have distinguished themselves by appearing frequently in previous active leads (e.g., statine, hydroxyethylamines, Freidinger lactams[16]). On the other hand, once a lead is available, most drug discovery proceeds through a series of evolutions (optimizations) in order to meet a set of predetermined criteria. Since specific structural types are sought, searching in a very broad pool of diversity (as above) is unlikely to be successful (actually it could uncover a new lead but is less likely to optimize an existing one). Ideally, what is required in this type of diversity-generating strategy is to "explode" around the known lead, i.e., to create as highly diverse population as possible that bears close structural resemblance to the original hit, followed by a selection for desired criteria.

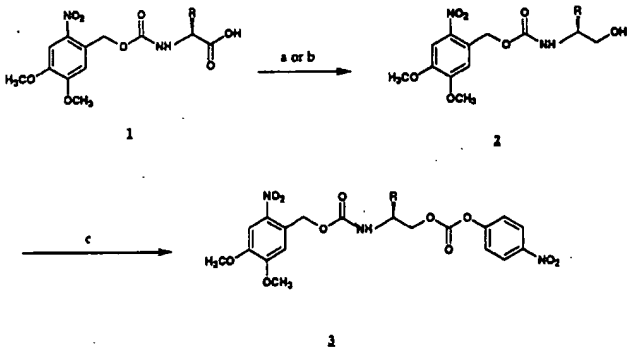
Clearly the subunits which lead to predetermined structures must be quite specific: from where should building blocks for known structural classes of pharmacophores arise? The answer, as in organic synthesis, lies in a retrosynthetic analysis or what we may term a *retrocombinatorial* approach to building-block selection. Lead structures should be retrosynthetically dissected in the maximum number of ways and upon these various possibilities imposed the needs of performing combinatorial chemistry. Inspection of the retrosynthetic tree invites the following key questions: By which modes of forward synthesis are the most building blocks available or obtainable? If the synthesis is allowed to proceed by that course, what is the scope and degree of reliability of the necessary reactions? Extending this line of reasoning should permit the maximum leverage to be applied combinatorially.

A common feature of both paradigms is likely to be a reliance on solid-phase-synthesis methods to facilitate the assembly of combinatorial libraries. Synthesis on a polymeric support greatly simplifies the problem of product isolation from reaction mixtures and also facilitates the partitioning of products into multiple aliquots for subsequent chemical elaboration. Moreover, the opportunity exists to take advantage of the support-tethered diversity in the design of convenient receptor binding assays for library evaluation. While there has been a long tradition of polymer-supported organic chemistry,[17-20] it is only in the areas of peptide and



Polycarbamates

Figure 2. Structure of a synthetic oligocarbamate prototype.



1 2 3

Figure 3. Synthesis of N-protected nitrophenyl carbonate monomers. Key: (a) BH_3, THF; (b) DCC, methylene chloride, *N*-hydroxysuccinimide, HOBt; then sodium borohydride, ethanol; (c) *p*-nitrophenyl chloroformate, methylene chloride, pyridine.

oligonucleotide synthesis that solid-supported chemistry has truly been optimized and become common-place. The advent of combinatorial organic synthesis will undoubtedly signal a renaissance in solid-phase organic chemistry, as workers attempt to adapt well-characterized homogeneous reactions to reliable solid-supported protocols.

Progress to Date: Synthetic Polymeric Diversity

The design and synthesis of novel synthetic monomers which, when assembled in a combinatorial fashion, could yield relatively low molecular weight polymeric materials is an approach that is well-suited to diversity generation and evaluation. Combinations of such monomers could lead to substances with novel backbones, possibly possessing desirable properties, such as metabolic stability, enhanced pharmacokinetic profiles, and cell and membrane permeability. Identification of these and other potentially modifiable parameters in such systems could facilitate drug discovery.

Schultz and co-workers have reported the synthesis of a library of oligocarbamates starting from a basis set of chiral aminocarbonates[21] (Figure 2). The monomeric units were readily obtained by the modification of amino acids via the intermediacy of the corresponding chiral amino alcohols (see Figure 3). The resulting nitrophenyl carbonate building blocks (*3*) were stable for several months at room temperature.

Oligocarbamates were synthesized on a solid support by deprotection of a resin-bound amine, protected with either the base-labile Fmoc or photolabile nitroveratryloxycarbonyl (Nvoc) group, followed by treatment with a nitrophenyl carbonate of type **3**. The deprotection/coupling cycle was repeated until an oligocarbamate of the desired length was attained (seven or eight cycles). Overall coupling yields were greater than 99% per step. Side-chain deprotection followed by resin cleavage afforded the desired oligocarbamates (Figure 4).

The VLSIPS photolithographic chip format, previously employed for oligopeptide synthesis, was used in the construction a spatially-addressable oligocarbamate library of 256 members. An anti-carbamate monoclonal antibody served as a model receptor for screening against this array. Antibody:oligocarbamate complexes were

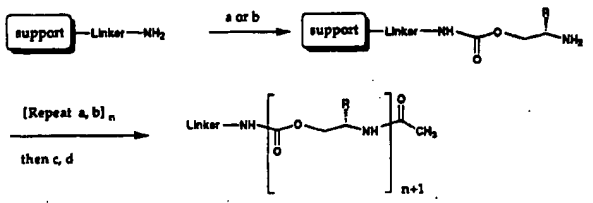


Figure 4. Solid-phase synthesis of oligocarbamates. Key: (a) nitrophenyl carbonate monomer, HOBt, diisopropylethylamine, NMP; (b) piperidine, NMP or *hν*; (c) acetic anhydride, NMP; (d) TFA, triethylsilane.

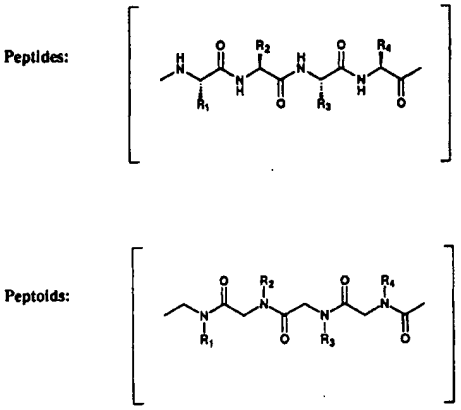


Figure 5. Comparison of peptide and peptoid backbones.

detected by treatment of the chip with a fluorescein-conjugated secondary antibody, followed by analysis using scanning epifluorescence microscopy. Because the location and structure of each different library member is defined by the synthetic strategy (binary masking) used in this technique, the necessity of sequencing the products is obviated. The binding activities of putative hits were confirmed by conventional assays using authentic material prepared by independent synthesis. A preliminary evaluation of the physicochemical properties of oligocarbamate molecules indicate that they are more hydrophobic than the corresponding peptide homologs, and their expected resistance to several proteases was confirmed.

Another type of synthetic polymeric diversity has been developed by Simon *et al.*[22,23] Through a variety of preparative routes, this group created a basis set of monomeric N-substituted glycine units, each bearing a nitrogen substituent similar to those of the natural α-amino acid side chains. The formal polymerization of these monomers results in a class of polymeric diversity which these workers have termed "peptoids" (Figure 5). Peptoids may be synthesized either manually or robotically following either a "full monomer" oligomer synthesis or via a "submonomer" synthesis, as reported by Zuckermann *et al.*[24] and illustrated in Figure 6. Various biological activities have been established for specific peptoid sequences, including inhibition of α-amylase and the hepatitis A virus 3C protease, binding to the tat RNA of HIV[22], and antagonism at the α_1-adrenergic receptor.[25] The peptoid approach to diversity generation has been extended to the preparation of encoded combinatorial libraries, in which natural amino acids code for the structure of the peptoid chain[26] (see part 1[1] and Figure 7).

An important variant of the synthetic polymeric diversity approach is directed toward construction of a chemical library in which the peptidyl backbone is conserved but a dipeptide unit is replaced at specific

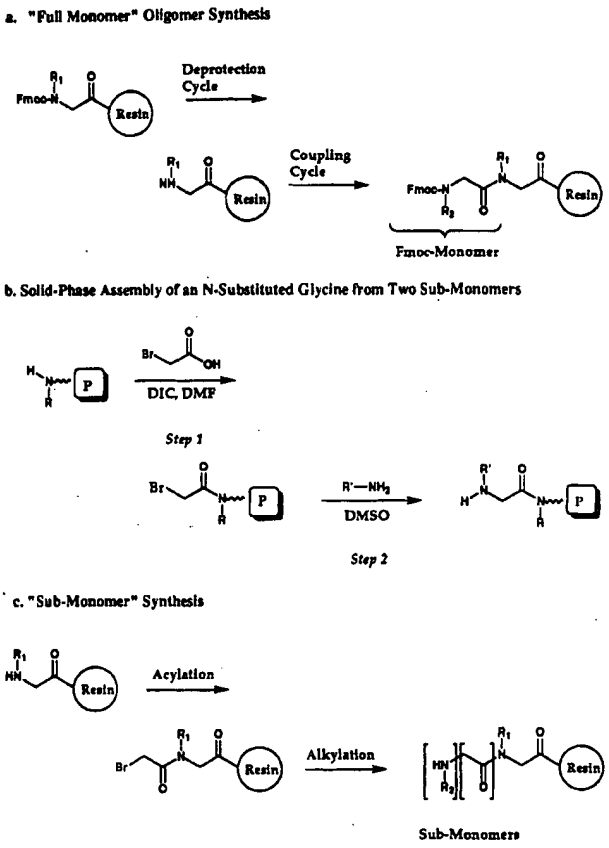


Figure 6. Synthesis of peptoids.

positions by a phosphonate dipeptide surrogate (see Figure 8). Such phosphonate pharmacophores are well-known as transition-state analogs for amide bond cleavage and have found wide usage in the inhibition of metalloproteases.[27-29] Campbell has described methods for the solid-phase synthesis of peptidylphosphonates that are compatible with the Fmoc/^{t}Bu protecting group strategy of standard peptide synthesis.[30] The key reaction step is formation of the phosphonate ester bond, which is achieved via a modified Mitsunobu condensation (Figure 9). Precursor lactic acid and protected amino phosphonate building blocks are prepared as shown in Figure 10.

When this process is applied to the combinatorial synthesis of peptidylphosphonates, the diversity product will be a metalloprotease enzyme inhibitor library. Enzyme-inhibitor libraries of this type and those focusing on other known inhibitory pharmacophores (thiols, hydroxamates, carboxyalkyldipeptides, etc.) may prove to be important tools in rapidly profiling novel proteases and for determining which pharmacophores are most effective at their inhibition. Using this knowledge, secondary inhibitory libraries can be constructed to optimize original leads. Through such a process it may be possible to dramatically accelerate the process of finding high-affinity enzyme-inhibitor ligands.

Another interesting type of polymeric diversity based upon a vinylogous polypeptide backbone has recently been reported by Hagihara *et al.*,[31] in which introduction of a trans olefinic linkage between the α-carbon and the carbonyl group of various amino acids is generalized. Additionally, Smith and colleagues have synthesized a non-amide polymer of (3,5)-linked pyrrolin-4-one oligomers which mimic the β-strand conformation of a normal peptide chain[32] (see Figure 11).

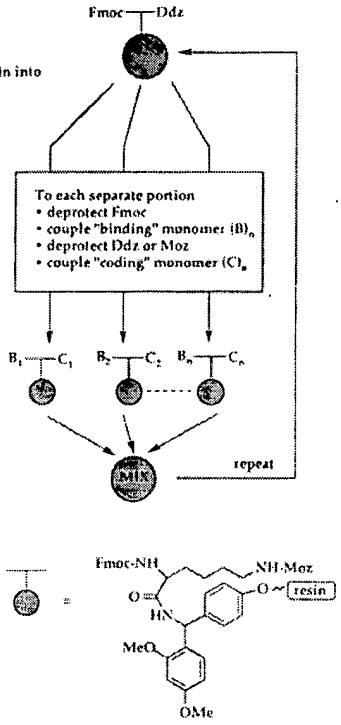


Figure 7. The synthesis of an encoded library consists of the following steps: (1) A bifunctional linker containing two orthogonally protected points of extension (Nα-Fmoc-Nε-Moz-Lys-OH) is attached to polystyrene resin via an acid-labile linker. (2) The solid support is divided into n equal portions at a mixture position. (3) A unique Nα-Fmoc-protected non-natural monomer (B) is coupled to the "binding" strand. (4) A series of Nα-Ddz-protected L-amino acids (C) are then coupled to the "coding" strand. (5) The solid supports are recombined.

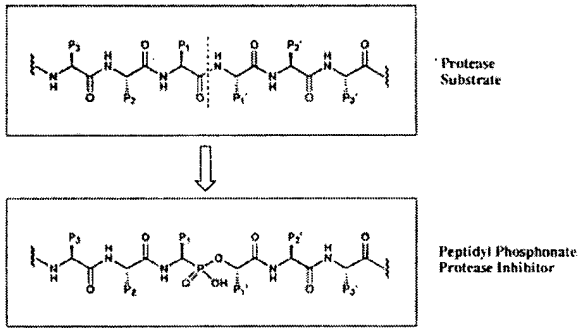


Figure 8. Peptidyl phosphonates.

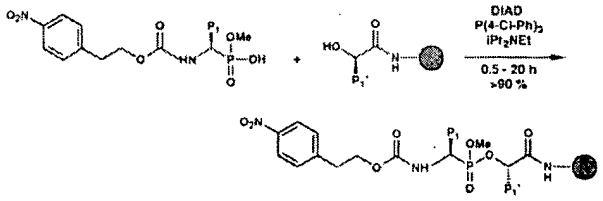


Figure 9. Solid-phase peptidyl phosphonate synthesis (SPPPS).

Nonpolymeric, Small-Molecule Diversity. The majority of chemical diversity generation discussed above concerns the preparation of linear molecules, in which the target structures are unambiguously specified by the order of building-block addition. In contrast, the great preponderance of organic synthesis proceeds rather differently, wherein building blocks interlock to give rise to

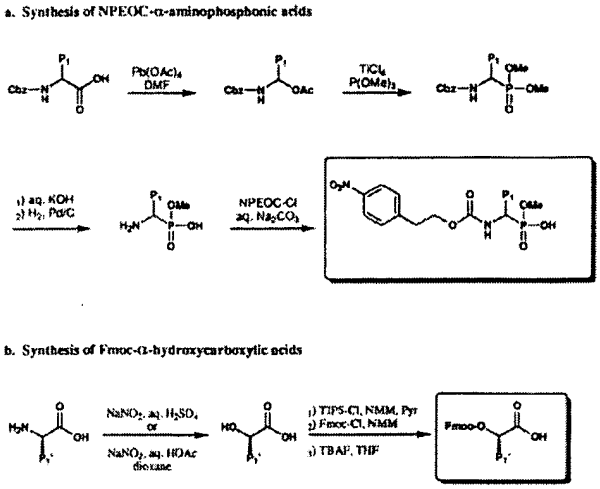


Figure 10. Synthesis of peptidyl phosphonate building blocks.

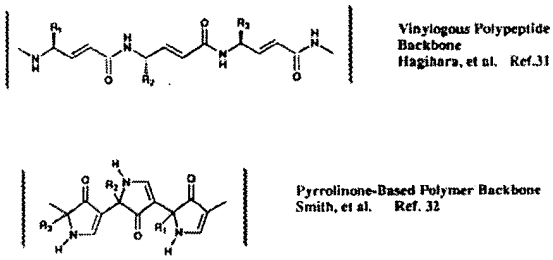


Figure 11. Novel polymeric backbones.

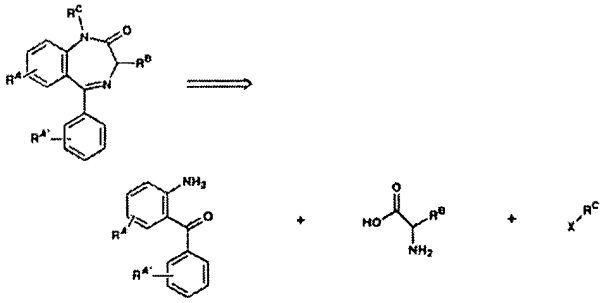


Figure 12. Components of a benzodiazepine library.

nonpolymeric, three-dimensional arrays. The recent seminal work of Ellman on the solid-phase synthesis of 1,4-benzodiazepines lays the groundwork for creation of a small-molecule library of one of medicinal chemistry's most notable pharmacophores and represents one of the first examples of the application of combinatorial organic synthesis to nonpolymeric organic compounds.[33]

The benzodiazepines are synthesized on a solid support by the connection of three building blocks, each of different chemical families (Figure 12). Following the attachment of 2-aminobenzophenone hydroxy or carboxy derivatives to the support using an acid-cleavable linker, [(4-hydroxymethyl)phenoxyacetic acid], the N-protecting group is deblocked (piperidine/DMF), and the weakly nucleophilic amine is acylated with an α-Fmoc-protected amino acid fluoride, using 4-methyl-2,6-di-*tert*-butylpyridine as an acid scavenger (Figure 13). Fmoc deprotection, followed by treatment with 5% acetic acid in DMF, causes the general cyclization to the intermediate lactam. Capitalizing on the ability of lithiated 5-(phenylmethyl)-2-oxazolidinone to selectively deprotonate the anilide NH,

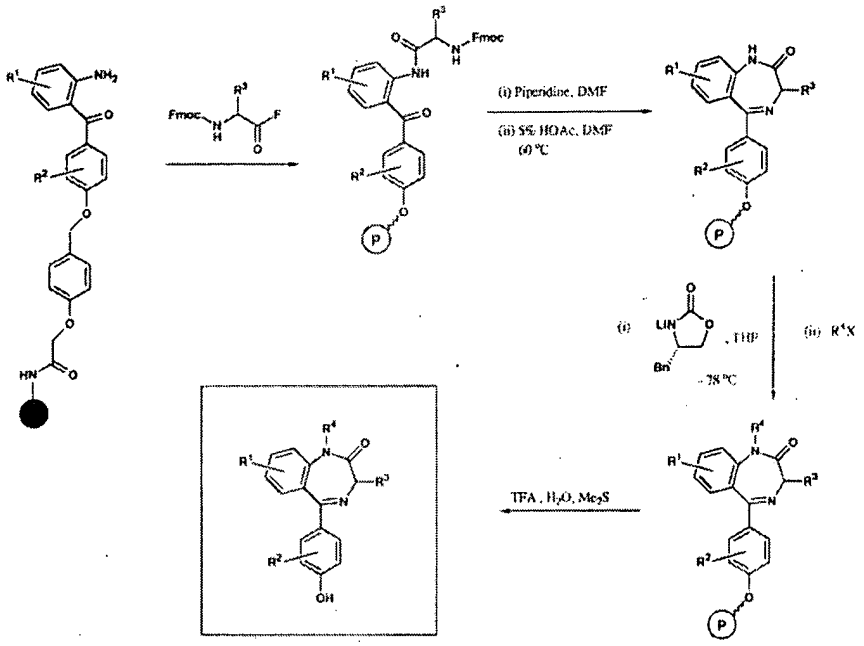


Figure 13. Combinatorial synthesis of the benzodiazepine pharmacophore.

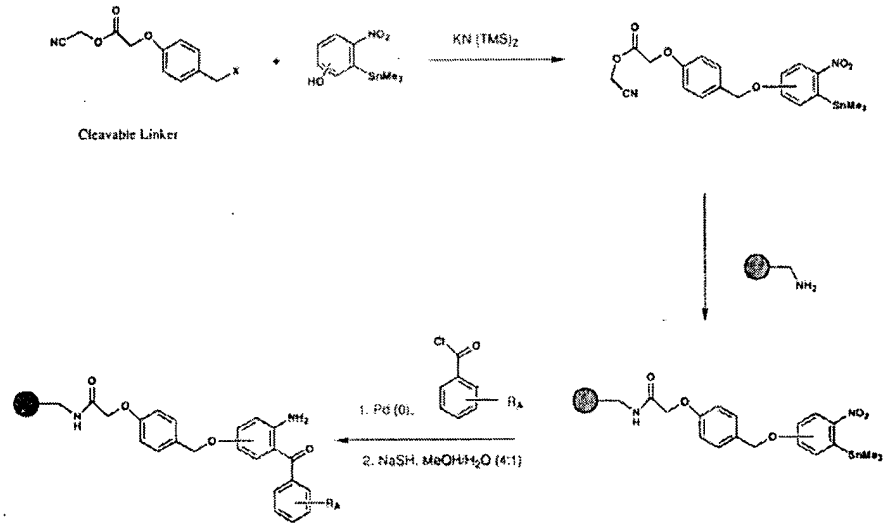


Figure 14. Synthesis of 2-aminobenzophenone derivatives on a solid support.

alkylation was achieved with a variety of alkylating agents. Aqueous acid cleaves the new benzodiazepine from the support in very high overall yields. The integrity of the chiral center was confirmed by a racemization test using chiral HPLC.

One of the limiting features of applying the above scheme to combinatorial library construction is that, though many alkylating agents and amino acid building blocks are commercially available, there is not a ready supply of appropriately functionalized 2-aminobenzophenones. Ellman addressed this problem directly by creating a general method for preparation of these materials on solid supports[34] (Figure 14). The stage is now set for the Ellman laboratory to create a benzodiazepine library.

Several other approaches to nonpolymeric molecular diversity have recently been published. In pursuit of small-molecule libraries, Nikolaiev *et al.* have used their amino acid encoding format (part 1[1]) with a building block basis set combining both amino acids and other synthetic units to prepare collections of nonpeptidic compounds and peptides refractory to Edman degradation (N-blocked peptides).[35] Representative examples of molecules which have emerged from such non-peptide libraries are shown in Figure 15.

A feature of several of the formats used in the display of synthetic diversity is that the potential ligands are tethered to a solid support. While screening strategies have been developed to exploit this feature, it is frequently desirable to screen compounds in solution. Many groups have engaged in developing releasable linker strategies to solubilize potential ligands. The issue has been addressed by a considerably different strategy by Hobbs DeWitt *et al.*, in which solid-phase chemistry, organic synthesis, and a designed parallel reaction apparatus were utilized for the generation of small-molecule libraries, the individuals of which, were termed "diversomers".[36] Target compounds which included dipeptides, hydantoins, and benzodiazepines were synthesized simultaneously but separately,

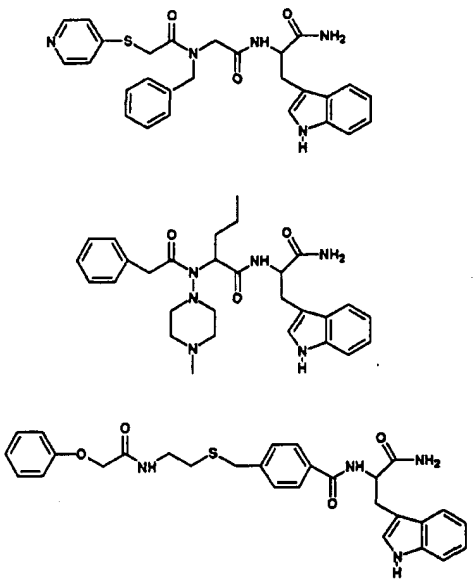


Figure 15. Structure of representative molecules from the Nikolaiev *et al.* nonpeptide library (ref 35).

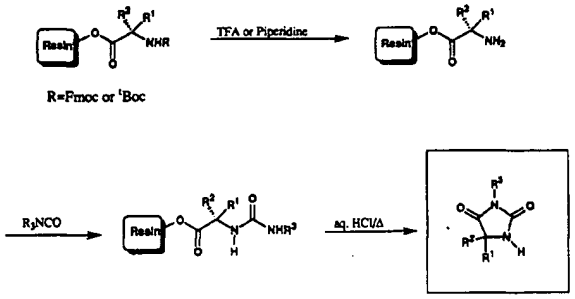


Figure 16. Synthesis of a hydantoin library.

on a solid support in an array format, to generate a collection of up to 40 discrete structurally related compounds. The preparation of hydantoins was carried out as shown in Figure 16. The synthetic strategy is directed through a resin-tethered penultimate product, in which the orchestrated revealing of distal functionality initiates attack on the resin-linking bond to eject the newly formed product into solution. Products which fail to react, should remain attached to the solid phase, and thus aid in product purification. The yields of hydantoins released from the resins in the final step ranged from 4 to 81% on a scale of 0.3–11 mg, which should be sufficient to support most preliminary *in vitro* biological testing. The resulting soluble, small molecules were characterized by traditional means. The authors also note the utility of ^{13}C gel-phase NMR to monitor reaction progress of the resin-bound intermediates.[37,38]

In a similar manner, a general method for multiple, simultaneous synthesis of soluble benzodiazepines was developed (Figure 17). Eight groups of five-amino acid resins were trans-imidated with five groups of eight 2-aminobenzophenone imines to form 40 resin-bound imines. Treatment with TFA liberated 40 discrete benzodiazepines from the resins. The products were obtained in 2–14-mg quantities, corresponding to 9–63% yields with estimated purities of >90%. Though the numbers of compounds involved in the diversomer methodology (~40) are significantly smaller than that which can be prepared by other library strategies (10^4–10^8), this interesting

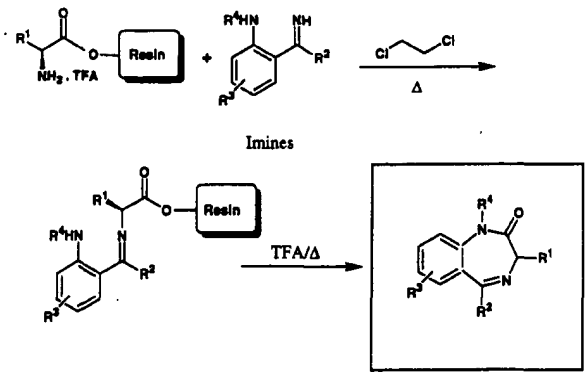


Figure 17. Preparation of a soluble benzodiazepine library.

approach to parallel organic synthesis produces relatively pure materials on a preparative scale in the traditional, soluble format.

Future Innovations

As the field of combinatorial chemistry receives increasing attention from the pharmaceutical establishment, it seems likely that the contents of chemical libraries will continue to evolve to look more and more like the type of compounds which have previously led to drugs. In spite of the complexity which early parts of the process may pose to the combinatorial chemist, a hidden advantage which combinatorially-derived molecules offer is that any "hit" will be readily synthesizable, by definition. This should be contrasted with a natural product driven approach to drug discovery and development, where often the structural complexity of the lead compound hampers the rapid preparation of analog molecules and the acquisition of SAR.

A previous point deserving further emphasis is that the vast universe of synthetic organic reactions are idiosyncratic transformations that fail to afford quantitative yields of unique products. Most synthetic chemistry procedures afford multiple products (regio- and stereoisomers, etc.) in variable yields. If diversity-generating chemistry proceeds ambiguously, how then are the results of small-molecule combinatorial organic syntheses to be understood and appropriate information extracted from library analysis? It may be speculated that encoding techniques will provide one method by which the combinatorial organic chemist can address the practical inefficiencies of chemical synthesis. Instead of envisioning an encoding tag as explicitly specifying the structure of an associated entity, one might consider the tag as *a record of the chemical history of individual library members*. Thus, after encoding the "recipe" or synthetic protocol used in the assembly of a combinatorial library, the library may be screened for active recipes. Once identified as "active", the synthesis would be replicated on a preparative scale, and the product mixture fractionated to identify active product(s). This strategy shifts emphasis from the criterion of singularity in a reaction outcome (a single predictable structure) to reproducibility and compatibility (orthogonality) with chemistry used in the synthesis of the encoding tag and in preparative scaling. The creation of encoded, small-molecule diversity, which can be released from a support (solubilized) while some type of link to the original tag is also maintained, is also likely to be an important area of investigation.

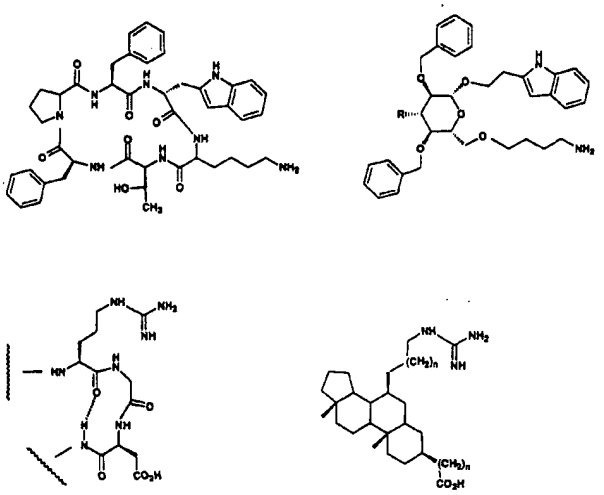


Figure 18. Scaffolds as templates for combinatorial chemistry.

One type of noteworthy chemical strategy which may have a bright future in the combinatorial realm rests on the conceptual extension of the work of Hirschmann, Nicolaou, and Smith and their co-workers into designing nonpeptidal peptidomimetics by the appropriate functionalization of designed scaffoldings (Figure 18). A specific example of this approach is the design and synthesis of a β-D-glucose-based nonpeptide mimetic of a potent cyclic hexapeptide somatostatin agonist.[39,40]

Hirschmann and co-workers have also used a functionalized steroidal template to serve as a backbone for mimicking a type 1 β turn.[41,42] The fact that such a remarkable job of molecular mimicry can be achieved with individual compounds bodes well for the application of this approach to combinatorial methods.

No library will ever be "complete" but instead will sample a subsection of a particular universe of molecular structure and space. In certain situations, libraries may be considered to be starting materials for the construction of new libraries of diversity. It is also useful to consider chemical libraries as collectable or archivable entities. Ideally, one seeks to preserve new compound libraries and use them for a variety of present and future screening needs. As time passes, the combinatorial chemist will be in possession of an accumulating collection of molecular diversity with which to challenge new drug targets. If chemical libraries are to become an item of commerce, a good deal more will have to be learned about their "shelf life" and how best to store them for future use. To date, the shelf life of chemical libraries is an open question.

B. Methods for Screening Combinatorial Libraries

The importance of distinguishing between the two principal applications of combinatorial technologies in ligand discovery, i.e., broad screening versus directed analoging, is particularly relevant to the design of assay methodologies for library evaluation. In searching large, highly diverse libraries for novel lead compounds, a premium is placed on the ability to detect rare ligands that may have modest affinity for the target receptor. The assay strategy may differ in screening analog libraries, since one is trying to develop quantitative SAR on a large number of compounds and to increase the potency of a lead. Regardless of the application, successful use of combinatorial libraries is highly dependent on the sensitivity and specificity of the assays that are used to identify and characterize ligands.

In this section, the various combinatorial library methods will be discussed in terms of the assays that are used. The assay formats are closely matched to the mode of presentation of the diversity. In broad terms, assay procedures can be grouped into three categories: (i) those that rely on affinity purification with an immobilized target receptor, (ii) those in which a soluble receptor binds to tethered ligands, and (iii) those in which soluble compounds are tested for activity, either directly or in competition assays. Each format presents different challenges with regard to the minimum affinity requirements for ligand detection, the demonstration of binding specificity, and the ability to discriminate among compounds in the library on the basis of their affinities for the target.

Isolation of Ligands by Affinity Purification

Recombinant Peptide Libraries. The various systems described in the first part of this series[1] for creating vast libraries of recombinant peptides (commonly referred to as peptide/nucleic acid complexes below) rely on affinity purification to select peptides that bind to a receptor. Two distinct methods have been employed to achieve affinity purification of peptide/nucleic acid complexes. The first involves incubation of a receptor in solution with the mixture of complexes. After allowing sufficient time for binding, the receptor is captured using immobilized streptavidin or an antireceptor antibody.[43,44] The second approach calls for preimmobilization of the receptor on beads, microtiter wells, or a chromatography support, followed by capture of the complexes.[45] In both cases, the use of a solid-support facilitates the separation and washing of receptor-bound complexes.

The method of receptor immobilization is a critical aspect of the successful use of recombinant peptide libraries. Because of the tremendous levels of ligand enrichment attainable through multiple rounds of selection and amplification, peptides that bind to any component of the solid support can be isolated from libraries. Peptides binding to streptavidin,[46,47] antireceptor antibodies,[48] or peptides that exhibit inherent nonspecific binding are readily isolated. Often, screening strategies employing subtractive methods and/or blocking ligands, are necessary to enhance the selection of ligands with desired binding specificity.

To enhance the probability of isolating peptide ligands with biological function, it is important to ensure that the receptor is active (for example, capable of binding its natural ligand) when immobilized. Immobilization of receptor proteins on microtiter wells or beads can be accomplished by passive adherence, covalent attachment, biotinylation and immobilization on streptavidin, or capture with high-affinity nonblocking antibodies. The first three processes often result in inactive proteins. The problem of immobilizing active protein can often be overcome by introducing into the receptor an immobilization handle through genetic engineering techniques. Peptide epitopes for a monoclonal antibody or a motif that allows for site-specific biotinylation of the protein[49] can be fused to proteins for this purpose. Generic immobilization strategies of this type greatly facilitate the creation of a high-density affinity matrix suitable for isolation of ligands.

Some drug-discovery targets may not be readily available as pure soluble receptors. Peptide ligands specific for the integrin IIb/IIIa have been successfully isolated from phage libraries by screening against platelets expressing a high density of this receptor.[50] It remains to be seen whether whole cells bearing receptors or other forms of impure macromolecular target will generally be successfully utilized to identify ligands. It can be anticipated that the problem of isolating non-receptor-binding sequences will be enhanced when dealing with impure forms of receptors.

The choice of using preimmobilized receptor rather than incubation with receptor in solution followed by receptor capture on a solid support may become important in one aspect of library screening. Because the phage and LacI/DNA complex systems are capable of displaying multiple copies of a peptide, multivalent binding can occur during affinity purification if receptors are immobilized at sufficient density. Multivalent binding effectively increases the avidity of the bound complexes and allows the isolation of complexes bearing peptides of lower affinity. Although it is possible that multivalent interactions may occur during the capture step of the two-step procedure, the stoichiometry of the phage or LacI complex, receptor, and the capture reagent need to be carefully controlled. Multivalent binding conditions may be more easily engineered with a high-density preimmobilized receptor.

The multivalency of the phage and LacI systems can be exploited to isolate peptides of modest affinities (K_d values of 1–1000 μM). This feature is particularly important in screening random libraries of peptides greater than six or seven amino acids in length. Libraries that can be routinely made have many fewer members (10^8–10^{10}) than the theoretical number of possible sequences for a given peptide length (for example, there are 1×10^{13} possible 10-mers). In general, peptide ligands for a receptor target consist of families of related sequences with few high affinity ligands in the family. Therefore, when a library is created, it is likely that only lower affinity members of the family will be represented. The ability to identify these low-affinity ligands then permits one to proceed to the secondary phase of discovery, the screening of mutagenesis libraries.

Methods for creating many variants of an initial sequence have been described in part 1. Such libraries will generally contain many low-affinity ligands and perhaps some high-affinity ligands in much lower abundance. A demand is therefore placed on the ability to selectively isolate the highest affinity ligands. Several methods have been described for efficient affinity selection. All are based on the prevention or disruption of multivalent interactions. The use of a low density of immobilized receptor to isolate high-affinity peptide ligands from a phage library containing many low-affinity ligands has been demonstrated.[44] Low receptor density reduces the possibility of multivalent interactions between phage particles and the receptor matrix. The "monovalent-phage" approach[45,51,52] has been successfully employed to isolate high-affinity mutants of human growth hormone displayed on phage. In this approach, phage particles with only a single chimeric pIII protein are created, thereby eliminating multivalent binding to immobilized receptor. Using this approach, mutants with K_d values of less than 5 pM have been identified.

An alternative method of affinity selection that may have advantages over other approaches has been developed.[48,53] Phage (or LacI particles) are allowed to bind in a multivalent fashion to a high density of immobilized receptor. For phage-bearing low-affinity ligands, the peptide on an individual pIII protein may be rapidly dissociating and reassociating, but the phage particle will not dissociate unless all the peptides on pIII are simultaneously in the unbound state. Dissociation of the phage can be initiated by addition of a competing ligand, which prevents rebinding of any individual peptide in the complex. Using a model system with peptides of known affinity, it was demonstrated that phage-bearing high-affinity peptides are retained for a greater length of time than phage with lower affinity sequences.[48] The concentration (and affinity) of the competing ligand, as well as the time and temperature of elution, can be varied to select for ligands of various affinities. This method has the advantage of using a high receptor density to ensure a full sampling of ligands of all affinity classes.

Achieving affinity selection is only part of the process of successfully screening recombinant peptide libraries. After selection, it is necessary to establish the binding specificity and, if possible, the affinity of individual peptides that result from the selection. Various assays have been described, including dot blots,[54] colony lifts,[55] and ELISA's with immobilized phage or immobilized receptor.[44] These methods differ in the minimum ligand affinity that is required for detection. In general, assays in which phage or LacI are immobilized (ELISA's, dot blots, colony lifts) require higher affinity (K_d values < 1 μM) peptides for detecting specific binding, and are therefore useful when such ligands are present in the selected pool. However, for reasons cited above, detection of the specific binding of lower affinity ligands is often necessary. In such cases, assays that use a high density of immobilized receptor are required to allow for multivalent binding and to increase the sensitivity of detection. If high-density receptor matrices are used for affinity purification and assay of individual clones, peptides with K_d values as high as 100–500 μM can be isolated with phage and LacI systems.[48]

An additional assay format has been described for estimating the affinity of peptides displayed by individually selected phage clones.[44] Radiolabeled receptor is first allowed to bind to the phage-borne peptides in solution. A high concentration of competing peptide is then added to prevent further binding, and the dissociation of radiolabeled receptor is followed with time. With a monoclonal antibody model system, a good correlation was observed between dissociation rates and the affinity of the peptide determined by solution methods. It has also been suggested that colony lifts with limited receptor concentration may allow discrimination of individual phage clones on the basis of their peptide's affinity.[55] However, this method may be confounded by differential levels of expression of phage by different colonies.

In summary, the successful identification of ligands from recombinant random peptide libraries depends not only on the nature and size of the libraries but also on effective screening strategies. Selection methods and assays of individual clones vary in their ability to select and detect lower affinity peptides and in the ease with which binding specificity can be determined. High-affinity ligands are most desirable, and initial conditions for screening of random libraries can be adjusted so that only high-affinity ligands are selected. However, for reasons stated above,

engineering selection and assay conditions to allow isolation and detection of lower affinity ligands may be generally a more reliable strategy. These initial peptides can serve as starting points for creating secondary recombinant peptide libraries or as leads for refinement by synthetic chemical combinatorial methods.

Affinity Purification of Mixtures of Soluble Synthetic Compounds. Investigators have employed affinity purification methods to isolate ligands from mixtures of soluble peptides[26,56,57] and oligonucleotides (RNA or DNA).[58-60] In the case of nucleic acid libraries, one takes advantage of the ability to enzymatically amplify the molecules resulting from affinity purification, and as with the recombinant peptide systems, multiple rounds of selection and amplification are used. Theoretical considerations in optimizing conditions for the selection of high-affinity oligonucleotides have been described.[61] The authors illustrate (by way of computer simulation) the importance of nucleic acid and receptor concentrations as well as the efficiency of separating specifically bound molecules. Computer simulations show that, under ideal conditions, rare high-affinity molecules can be isolated from large libraries with relatively few rounds of selection and amplification. There have been a number of examples of successful identification of high-affinity oligonucleotides using this process.[62]

Relatively little work has focused on the affinity purification of ligands from soluble peptide libraries. The isolation of ligands for an anti-gp120 antibody from equimolar mixtures containing 19 or 32 peptides has been reported.[56,57] More recently, the same antibody was used to capture ligands from four mixtures, each of 50 peptides, comprised of unnatural amino acids fused to an encoding L-amino acid peptide strand.[26] After affinity purification, the resulting pool of peptides selected was resolved by HPLC and each peak subjected to Edman sequencing and mass spectrometry analysis. A major limitation of this approach is the sensitivity of these analytical methods. Sufficient peptide (>1–10 pmol) must be recovered in order to determine its sequence, requiring that each library member be present in relatively high amounts in the starting pool and that there be sufficient receptor available to isolate the requisite quantity of each of the high-affinity ligands. In addition, the method requires that selected peptide ligands be resolved chromatographically. While it seems unlikely that this methodology will be extended beyond libraries of modest size (less than a few thousand members), it may prove a useful technique for evaluating secondary (analog) libraries. The proposed approach of creating a library of soluble compounds with attached oligonucleotides tags may allow for the structural identification of minute quantities of compounds isolated by affinity purification.[63,64]

In theory, chromatography of compound mixtures using receptor columns should not only facilitate separation of nonbinding members of the library, but should also allow for the resolution of compounds on the basis of their receptor affinities. Work with various model systems has demonstrated that column retention time can be used as an index of affinity.[65,66] While columns of receptor target have been used in batch affinity purification methods, chromatography to resolve ligands of differing affinities has yet to be applied to screening combinatorial libraries. This method may be better suited to isolation of ligands of moderate affinity.[67] An additional limiting factor in the use of chromatography may be the large amount of receptor required to generate enough theoretical plates to effectively resolve compounds.

Binding of Receptors to Immobilized Ligands

Various methods for creating libraries of compounds attached to solid supports (pins, beads, chips, etc.) have been outlined in part 1.[1] Such libraries are screened by detecting the direct binding of a labeled receptor to an immobilized ligand. The identify of the ligand is either determined directly (by peptide sequencing or mass spectrometry), specified by its spatial location in an array, or deduced by reading an encoding tag.

There are a number of important issues related to solid-phase binding assays with immobilized ligands. First, the ability of a receptor to interact with a tethered ligand may be influenced by the site or nature of its covalent attachment to the solid support. In all of the methods published to date, peptide ligands are attached to a linker and support via the carboxy terminus of the sequence. An obvious example of the limitation imposed by this mode of immobilization would be in screening against the G-protein-linked receptors of various peptide hormones, many of which require a free C-terminal carboxamide for activity. In such a case, it is likely that many peptide analogs that would bind when free in solution would be missed in an assay where the same peptides were immobilized via their C-termini. To circumvent this problem, it is advantageous to have several alternative sites of ligand attachment to the surface. Methods for tethering peptides through their N-termini have been identified.[68] It is likely that the issue of how best to tether molecules to surfaces will become even more important when dealing with libraries of small nonpolymeric organic compounds.

The chemical nature of the linkage between the ligand and support may also affect the receptor-ligand interaction. One needs only to look at the variety of resins that are available for affinity chromatography to appreciate the importance of controlling the receptor–ligand interface. The types of linker groups that have been successfully employed in tethered library assays to date have been noted in part 1 of the series. Whether these linkers will generally provide for optimal presentation of compounds to other receptor systems remains to be seen.

Immobilized ligand assays require that the receptor be labeled in a way that allows for highly sensitive detection of receptor binding. The receptor can either be labeled directly or a secondary labeled reagent with high affinity for the receptor can be used. To date, colorimetric enzymes, radioisotopes, and fluorophores have been used in labeling receptors or secondary reagents. The reagents must be labeled in a way that maintains the activity of the receptor, for instance, its ability to bind a natural ligand. This can be greatly facilitated by creating chimeric recombinant receptors that incorporate peptide epitopes of antibodies or peptide sequences for site-specific radioactive phosphorylation[69] or site-specific biotinylation.[49]

Successful screening of libraries of immobilized synthetic ligands is dependent on the same types of issues as have been previously discussed with respect to evaluating recombinant peptide libraries: i.e., the affinity threshold for detection, the ability to discriminate ligands on the basis of their affinities, and the ability to distinguish specific binding from nonspecific binding. Methods development in this area is in its infancy. In principle, it

should be possible to exploit multivalent binding to detect lower affinity ligands. Multivalent receptors can be created by a number of methods, including genetic fusions to generate bivalent receptor/Fc fusions[70] or through the use of monoclonal antibodies or streptavidin to create cross-linked receptors capable of interaction with more than one immobilized ligand. Optimization of the density of immobilized ligands may be required in order to allow for multivalent binding. As has already been noted, it may be important to be able to isolate relatively low-affinity ligands in the initial screening of random libraries. These compounds can then serve as the basis for further library construction in which the goal is to improve ligand affinity.

Affinity discrimination during the screening of either primary random libraries or secondary (analog) libraries is of obvious importance. There has been little published work on methodology in this area. In principle, low receptor concentrations, competing ligand-mediated dissociation, and/or stringent washing conditions can be utilized to identify the highest affinity ligands. Two issues complicate the use of such methods. The first is the likelihood that each pin, bead, or surface synthesis site does not contain the same amount of compound. With different compound loadings, one must be extremely cautious of using the absolute quantity of bound receptor as an index of a molecule's affinity. As new building-block and coupling chemistries are adapted to combinatorial formats, this may become a more significant problem than it is for high-yielding peptide chemistry. Another complicating feature of the immobilized ligand assay format is the fact that ligands of one particular kind are densely clustered on a surface. Both the association and dissociation rate constants of a receptor/ligand interaction are affected by surface ligand density. The binding of nearby ligands depletes the local receptor concentration and the association kinetics become diffusion limited. After dissociation, receptor rebinding is favored because of the high local-ligand concentration and the apparent dissociation rate is reduced. Theoretical and experimental analyses of these surface binding effects have been undertaken.[71] The impact that these surface binding kinetics will have on the ability to discriminate among library members on pins, beads, or glass surfaces remains to be seen.

The information generated by screening immobilized ligand libraries differs among the various library formats. In the case of bead-based technologies, compounds exceeding a threshold affinity are sampled from a large pool of ligands. Positive information is obtained, i.e., that a particular ligand binds to the receptor. One cannot, however, draw conclusions about the binding affinity of nonselected ligands. The sampling size may not have been large enough to include all high-affinity ligands, or a high-affinity bead may have been missed by the affinity selection method [for example, fluorescence-activated cell sorter (FACS) selection]. By contrast, the multipin and VLSIPS technologies allow one to perform a parallel assay in which data is obtained on every compound that is synthesized. In principle, both positive and negative binding information can be exploited in the design of second-generation compounds.

Incorporation of methods that assess the specificity of binding of ligands is an important aspect of screening random libraries. Screening immobilized ligands by direct receptor binding can lead not only to the identification of ligands of interest (for instance, ligands that compete with the natural ligand) but also to ligands that bind to undesired portions of the receptor or to secondary detection reagents. In the case of libraries of compounds on beads, it may be possible to remove undesired ligands in a subtraction step prior to screening for desired ligands. For compounds on pins or chips, it may be possible to make replicate arrays and test for total binding and nonspecific binding in parallel. Otherwise, sequential assays that first test for receptor binding of any kind, followed by an assessment of nonspecific binding will be required in order to correctly identify compounds that interact with the receptor in a desired manner.

Testing the Activity of Libraries of Soluble Compounds

The classical method of screening for a desired biological activity is to test soluble compounds one at a time in a competition binding assay, enzyme inhibition assay, or in a cell-based bioassay. Such approaches have been applied to library screening by releasing compounds synthesized on pins into microtiter wells, as described in section C of part 1 of this series.[1] A novel application of bead technology has recently been disclosed where compounds on individual beads are released locally onto a lawn of confluent mammalian cells and cause activation of cells in the area surrounding the bead.[72] The bead responsible for cell activation is isolated and a small amount of noncleaved peptide is sequenced to determine its structure. In both of these cases, the principal issue is whether enough compound is released to be detected by the assay. For pins, approximately 100 nmol of peptide can be released into a few hundred microliters of solution, while beads with diameters of $\sim$100 μm can release on the order of 100 pmol of peptide.

Rather than assaying compounds individually, a second approach to screening soluble libraries is to assay compound mixtures. In addition to testing complex pools of soluble peptides (*vide supra*), libraries of oligonucleotides have been successfully screened as soluble mixtures.[73] The most frequently used strategy for screening mixtures of soluble compounds with the goal of ultimately identifying single active molecules is based on the "mimotope" approach, detailed in part 1. The essence of this strategy is that degenerate pools of peptides (or other compounds) are resolved into their most active constituents by an iterative process of testing and resynthesis until a single sequence is identified as having high activity. A variation of the methodology (termed "bogus-coin strategy") has also been described.[74]

There are a number of caveats to using this methodology for testing soluble compound mixtures. In practice, the results of each set of assays do not typically indicate a preference for a unique residue at any position within the sequence. Rather, comparable assay results may be obtained for several different amino acid substitutions and some decision must be made as to which of these partial solutions should be fully resolved. The number of peptide mixtures to be synthesized and tested in this protocol expands dramatically as the number of alternative sequences selected for complete resolution at each cycle is increased. Moreover, the deconvolution of different partial solutions may frequently produce divergent resolved sequences, in part because the contribution of each amino acid to the peptide–receptor interaction is typically

dependent on other non-neighboring residues within the ligand. The problem of identifying the most potent ligand in a complex mixture by this type of iterative pathway is exacerbated by the relative abundance of lower affinity ligands that represent local binding optima.

Originally designed for identifying antibody ligands, the mimotope strategy has primarily been used for libraries of six to eight building blocks in length. It is not clear that ligands of this size will be optimal for other types of receptors (although success with opioid receptors[75] and other targets have been reported). As the length of the compounds in the library increases, resynthesis and testing of pools becomes more cumbersome.

Perhaps the greatest limitation of this methodology is the fact that the activity of a given pool is based on the cumulative activity of all the compounds in the pool; i.e., pools with the same activity may contain many low-affinity compounds or a few high-affinity compounds. For this reason, the methodology is greatly facilitated if the minimal fragment having activity is comprised of the same number of building blocks as used in constructing each library member (e.g., a uniquely active tetramer is more easily resolved from a tetrapeptide library than a hexamer library). Alternatively, the identification of active peptide(s) is facilitated if the receptor has specific requirements for a fixed position within a peptide ligand (e.g., the N or C termini). If neither of these conditions is true, it may be necessary to test many or all of the possible initial pools with two adjacent or nonadjacent fixed residues. This drastically increases the number of initial pools that need to be synthesized but increases the probability that a critical residue(s) is fixed in at least one pool to allow that pool to differentiate itself. It must be kept in mind that any pool identified as having the greatest activity may be composed of many moderately active compounds and that the most active compound(s) may reside in other pools.

Testing of mixtures of soluble compounds is also limited by the concentration of individual test compounds that can be achieved in the initial pools. Pools containing as many as 160 000 different peptides have been tested with each member being present at ~10 nanomolar concentration.[76] Because of limitations on the solubility of the total pool, the concentrations of individual compounds present in increasingly larger libraries must be correspondingly diminished. This will ultimately limit the ability to identify the activity of compounds with modest potencies.

While the current methods for testing mixtures of soluble compounds have certain drawbacks, screening soluble libraries does have the decided advantage of avoiding the problems associated with assaying tethered molecules in other combinatorial technologies. Conventional binding and enzyme and cell-based assays (including those with poorly defined biochemical targets) can be used to test the activity of soluble compounds. It is likely that in the future, encoding strategies will be employed to allow more facile screening of soluble molecules. In the simplest format, single encoded beads can be dispensed into microtiter wells. The compounds can then be released from the beads and tested for activity, with the identity of the most active compound(s) being deduced by decoding the tag attached to the bead(s). To test large libraries of soluble compounds, mixed pools of encoded beads can be created. At each round of testing, only a fraction of the compound is cleaved from each bead. Active pools of beads are pursued by further subdividing the beads, partially

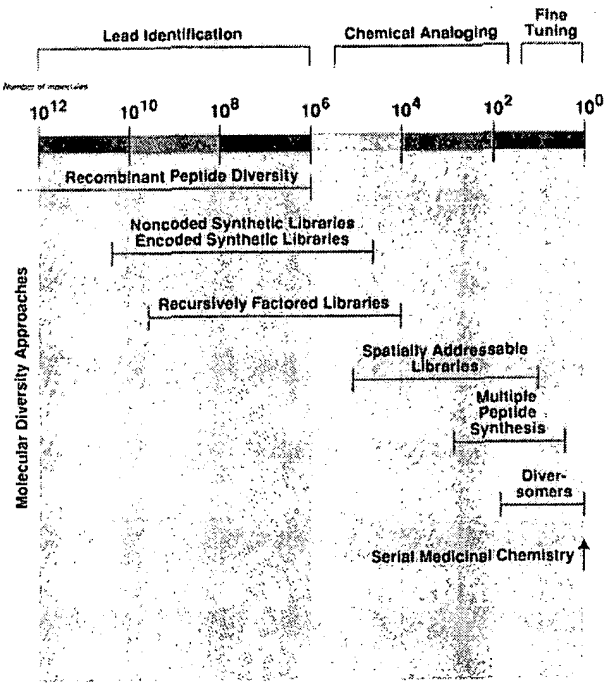


Figure 19.

releasing the compounds, and testing for activity. In the end, a single bead is identified with the greatest activity and the encoding tag is then read.[77]

C. Integration of Combinatorial Technologies for Drug Discovery and Future Directions of the Field

Regardless of whether the objective is a broad discovery search or analoging a known lead, a key aspect in the successful application of combinatorial technologies to drug discovery is the requirement for having a closely linked, coordinated process for the integration of synthesis and screening. The creation and evaluation of molecular diversity are two sides of the same coin. It is still premature to speculate on which type of format will prove most suitable for a particular type of discovery/development problem. Over time, a cataloging of successes and failures will clarify this issue. In all likelihood, command of a collection of combinatorial tools will be required for general success. One may imagine a "spectrum of molecular diversity" stretching from few to many molecules (see Figure 19). Serial synthesis (contemporary medicinal chemistry) operates on a few molecules (far right of chart). We expect that each individual combinatorial tool/format will turn out to be most valuable within proscribed regions of numbers of molecules. Hence, recombinant peptide diversity is particularly suitable for generating and screening large libraries of $>10^8$ compounds. The VLSIPS chip technology, although capable of displaying vast arrays, is primarily an analoging tool and is most useful for evaluating 10^2–10^4 compounds. The parallel organic synthesis methods of Hobbs DeWitt *et al.*[36] are applicable for tens of compounds. Encoded synthetic libraries appear to be relevant in broad diversity searching and may also prove useful in narrower optimization strategies. Given the repetitive nature of many of the manipulations required for library construction, an on-going priority will be to address the possibility of automating as many aspects of the generation/evaluation process as possible. Growth in

library size also impacts directly on the physical size of compound collections and on the amounts of target receptor required for screening. These pressures will continue to drive the field toward minaturization and exploitation of nanotechnologies.

The power of combinatorial technologies in generating huge numbers of compounds suggests that in a lead-discovery mode, less preconceived bias need be brought to the process of making molecules. Another way of expressing this is as follows: let the numbers do the talking. Due to the time and effort required in serial approaches, each target molecule is selected with great care. Because of the relative ease in creating libraries, little risk is incurred, or effort expended, in allowing a wide variety of building blocks to participate in diversity generation. Since there is less up-front investment in any individual combinatorially created molecule, the combinatorial chemist can afford to take more risks. We can think in terms of a portfolio of libraries which might be routinely applied to the initiation of a drug-discovery search. This is not dissimilar to selecting a preliminary screening sample of diversity from a large database of individual molecules.

A related, but still immature, issue in combinatorial approaches to drug discovery revolves around the idea of "quantitation of diversity". An understanding of the concept of "measuring" molecular diversity could impact on designing libraries to contain maximal structural diversity. This notion has arisen previously in deciding which few representative, highly diverse compounds to select out of large database collections, when setting up groups of preliminary screening samples. The huge numbers involved in combinatorial approaches intensifies this issue. A number of interesting approaches to the diversity quantitation problem can be expected to emerge.

One working drug-discovery paradigm might be based on initially employing a portfolio of biological diversity (peptide libraries) together with standard chemical libraries (various-sized cyclic peptides, cysteinyl-linked cyclics, etc.), peptides with carboxyl or carboxamide display, synthetic polymeric diversity, as well as large libraries of semirigid and acyclic small molecules prepared by COS. Over time, favored libraries and directions would be expected to emerge. As the sophistication of combinatorial organic synthesis grows, the origins of a molecular structure as either "combinatorially or serially derived" will gradually become transparent.

Another area where considerable effort must be applied is in the registry of libraries and individual library members. It is unclear that library compounds should be registered and documented for testing in the same ways as serially produced compounds historically have been, but exactly what changes are necessary remain to be determined. Vast numbers of compounds have been and are being created; keeping track of these and their corresponding biological activities will require innovative database-management techniques. Additionally, nomenclature needs to be developed by which one can simply express the constitution, scope, and nature of chemical libraries. Legal issues, including the patenting and documentating of libraries and their component members, will need to be pioneered.

As repeatedly emphasized, it is obvious but imperative to have efficient means of evaluating the molecular diversity which is generated. Different assay techniques will be format specific. Assays must clearly discriminate specific from nonspecific binding. Since in a broad screening mode, one is almost always sampling a small percentage of the entire universe of diversity (10^{10} peptides are only 0.1% of the universe of 10-mers), it is crucial that appropriate assay techniques be competent to detect modest affinity ligands. The identification of weak binders in any of the aforementioned approaches is very important and should lead directly to preparation of secondary libraries in which original "hits" will become the centerpiece for more focused diversity creation. This is a consequential issue, since application of combinatorial technologies are best viewed as an iterative process and not a singular event. As the emphasis shifts to analog evaluation, assays must be capable of affinity discrimination between closely related library members. The tools of molecular biology have permitted the molecular engineering of targets to serve the purposes of screening. The rapid introduction of targets into a screening mode will require generic techniques for their handling, and manipulation of molecular targets by genetic engineering will continue to play a crucial role in marrying library evaluation and synthesis. Though combinatorial technologies may soon prove their worth in the drug-discovery process by delivering new leads quickly and cheaply, in order to completely fulfill the promise of "making drugs", an important question will be whether some of the common major obstacles to drug development (e.g., cell penetration, bioavailability, pharmacokinetics, metabolism) can be productively addressed through the application of combinatorial approaches (i.e., *combinatorial drug development*).

In the coming years, cloning and sequencing of the human genome promises that an unprecedented abundance of newly discovered proteins will become available as potential drug targets. Gaining even more prominence than it now assumes will be the issue of discriminating among a myriad of receptors and enzymes to identify valid targets for drug discovery. The ability to access potent and specific ligands for these targets will guide this process by untangling the physiological relevance of endogenous biochemical pathways. Combinatorial methods will be called upon to provide such molecules to quickly and cheaply drive target validation. In this manner, the identification of leads will benefit from a significant, but hidden, benefit which emerges from combinatorial screening; hits derived from chemical libraries should be readily amenable to combinatorial analoging.

Certain drug targets may present more or less of a historical precedent with respect to the likelihood of successfully identifying a tight binding ligand through the use of known pharmacophores. For example, the search for specific enzyme inhibitors may be facilitated by the intentional enrichment in the combinatorial synthetic process of building blocks containing known inhibitory pharmacophores. Particularly important or common types of drug targets may justify having on hand special libraries which are somewhat specific (i.e., a peptidyl hydroxyethylamine library for aspartyl- and metalloprotease inhibition[78-80]). On the other hand, in areas where there is less current information (e.g., antagonism of protein–protein or carbohydrate–protein interactions), a wider scope of diversity search should be taken until consistent patterns begin to emerge. In the case of newer, less explored target groups, combinatorial technologies can be expected to assist in unearthing new pharmacophore

classes and to help establish an understanding of drug design for new types of targets.

Combinatorial technologies diverge sharply from historical precedent through a change in emphasis from the consideration of individual molecules to thinking in terms of populations of molecules. A common, but false, intuitive believe is that combinatorial chemistry is necessarily a random, screening search; the antithesis of rational drug design. In fact, all libraries are biased in some ways. All drug company compound files are biased by the historical programs of that institution, since a disproportionate share of compounds of particular types will have been deposited. *The notion of intentionally biasing a chemical library is a form of drug design*, but again not applied to individuals but rather to groups or populations of molecules. If a scientist hypothesizes on the basis of structural information that the current lead molecule contains a type II β-turn motif, then rather than performing two or three serial tests of this idea, the combinatorial chemist might create a library of narrow diversity utilizing a basis set of β-turn mimetics and thus interrogate many slightly different regions of conformational space simultaneously. The drug design of populations versus individuals is analogous to fishing with a net rather than just a hook. As more knowledge of workable strategies for combinatorial synthesis are understood, it is expected that structural and computational input and other rational design information will be integrated into a broad combinatorial medicinal chemistry approach.

Gaining a full appreciation of the issues and difficulties which must be surmounted in order to perform useful combinatorial organic synthesis will initially be a relatively slow process, especially because the important strategies and decision points differ so markedly from traditional organic synthesis. Retrocombinatorial analysis of existing pharmacophores and other important structures should assist in decision making; both in choosing routes of forward synthesis and in synthetic target selection. If combinatorial techniques are indeed to become a useful shortcut to new leads and optimized compounds, then one key implied goal of combinatorial organic synthesis is to intersect the pathway of modern medicinal chemistry upon which compounds move from the early discovery stages to clinical candidacy. Rich incentives await those who are able to mass produce important biologically active molecules quickly and cheaply. Not surprisingly, an aggressive, worldwide effort to understand and master this field has already begun.

This Perspective has been restricted to a consideration of the impact of combinatorial technologies on medicinal chemistry/drug discovery and development. From the point of view of applicability of the technologies, this is an artificially narrow view. Combinatorial processes will become important in diagnostic medicine,[81] agricultural chemistry, food chemistry, immunology, molecular biology, polymer studies, inorganic synthesis, and many other fields. Though the field of "combinatorial chemistry" is chronologically a new enterprise, the evolution of thought in this fertile area continues to outrace the experimental reduction to practice of many ideas. One may reasonably ask "why are combinatorial technologies happening now?". The answer is probably complex and beyond the scope of this Perspective. Nevertheless, the explosive recent interest in the application of combinatorial technologies to drug discovery is symptomatic of an idea whose time has come. Because the issues which confront the medicinal chemist differ so radically from historical approaches, the combinatorial field will no doubt continue to provide impetus and stimulation for the formulation of new concepts and ideas.

Acknowledgment. We thank Drs. A. Zaffaroni and G. Ringold for their interest, insights, and support in undertaking the application of combinatorial technologies to drug discovery and Mrs. Terri Nebozuk for help in preparing this manuscript.

References

(1) Part 1: Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries. *J. Med. Chem.* **1994**, *37*, 1233–1251.
(2) Barker, P. L.; Bullens, S.; Bunting, S.; Burdick, D. J.; Chan, K. S.; Deisher, T.; Eigenbrot, C.; Gadek, T. R.; Lipari, M. T.; *et al.* Cyclic RGD Peptide Analogues as Antiplatlet Antithrombotics. *J. Med. Chem.* **1992**, *35*, 2040–2048.
(3) Al-Obeidi, F.; De L Castrucci, A. M.; Hadley, M. E.; Hruby, V. J. Potent and Prolonged Acting Cyclic Lactam Analogues of an α-Melanotropin: Design Based on Molecular Dynamics. *J. Med. Chem.* **1989**, *32*, 2555–2561.
(4) Wiley, R. A.; Rich, D. H. Peptidomimetics Derived from Natural Products. *Med. Res. Rev.* **1993**, *13*, 327–384.
(5) Giannis, A.; Kolter, T. Peptidomimetics for Receptor Ligands—Discovery, Development, and Medicinal Perspectives. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1244–1267.
(6) Kates, S. A.; Sole, N. A.; Johnson, C. R.; Hudson, D.; Barany, G.; Albericio, F. A Novel, Convenient, Three-Dimensional Orthogonal Strategy for Solid Phase Synthesis of Cyclic Peptides. *Tetrahedron Lett.* **1993**, *34*, 1549–1552.
(7) Polinsky, A.; Cooney, M. G.; Toy-Palmer, A.; Osapay, G.; Goodman, M. Synthesis and Conformational Properties of the Lanthionine Bridged Opioid Peptide [D-Ala_2,Ala_5]Enkephalin as determined by NMR and Computer Simulations. *J. Med. Chem.* **1992**, *35*, 4185–4194.
(8) Albericio, F.; Hammer, R. P.; Garcia-Echeverria, C.; Molins, M. A.; Chang, J. L.; Munson, M. C.; Pons, M.; Giralt, E.; Barany, G. Cyclization of Disulfide-Containing Peptides in Solid Phase Synthesis. *Int. J. Pept. Protein Res.* **1991**, *37*, 402–413.
(9) Marlowe, C. K. Peptide Cyclization on TFA Lable Resin Using the Trimethylsilyl (TMSE) Ester as an Orthogonal Protecting Group. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 437–440.
(10) Trzeciak, A.; Bannwarth, W. Synthesis of 'Head to Tail' Cyclized Peptides on Solid Support by FMOC Chemistry. *Tetrahedron Lett.* **1992**, *33*, 4557–4560.
(11) Andrews, P. C.; Boyd, J.; Ogorzalek Loo, R.; Zhao, R.; Zhu, C.-Q.; Grant, K.; Williams, S. Synthesis of Uniform Peptide Libraries and Methods for Physico-Chemical Analysis. Poster presented at the Meeting of the Protein Society, San Diego, July 1993.
(12) Metzger, J. W.; Wiesmuller, K.-H.; Gnau, V.; Brunjes, J.; Jung, G.; Ion-Spray Mass Spectrometry and High Performance Liquid Chromatography–Mass Spectrometry of Synthetic Peptide Libraries. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 894–896.
(13) Zuckermann, R. N.; Kerr, J. M.; Siani, M. A.; Banville, S. C. Design, Construction and Application of a Fully Automated Equimolar Peptide Mixture Synthesizer. *Int. J. Pept. Protein Res.* **1992**, *40*, 497–506.
(14) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. Light-Directed, Spatially Addressable Parallel Chemical Synthesis. *Science* **1991**, *251*, 767–773.
(15) Zuckermann, R. N.; Kerr, J. M.; Siani, M. A.; Banville, S. C.; Santi, D. V. Identification of Highest-Affinity Ligands by Affinity Selection from Equimolar Peptide Mixtures Generated by Robotic Synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4505–4509.
(16) Freidinger, R. M.; Perlow, D. S.; Veber, D. F. Protected Lactam-Bridged Dipeptides for Use as Conformational Constraints in Peptides. *J. Org. Chem.* **1982**, *47*, 104–109.
(17) Crowley, J. I.; Rapoport, H. Solid-Phase Organic Synthesis: Novelty or Fundamental Concept. *Acc. Chem. Res.* **1976**, *9*, 135–144.
(18) Leznoff, C. C. The Use of Insoluble Polymer Supports in General Organic Synthesis. *Acc. Chem. Res.* **1978**, *11*, 327.
(19) Frechet, J. M. J. Synthesis and Applications of Organic Polymers as Supports and Protecting Groups. *Tetrahedron* **1981**, *37*, 663–683.
(20) Hodge, P. Organic Reactions Using Polymer-Supported Catalysts, Reagents or Substrates. In *Syntheses and Separations Using Functional Polymers*; Sherrington, D. C., Hodge, P., Eds.; John Wiley and Sons: New York, 1988, pp 43–122.
(21) Cho, C. Y.; Moran, E. J.; Cherry, S. R.; Stephans, J. C.; Fodor, S. P. A.; Adams, C. L.; Sundaram, A.; Jacobs, J. W.; Schultz, P. G. An Unnatural Biopolymer. *Science* **1993**, *261*, 1303–5.

(22) Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C. K.; Spellmeyer, D. C.; Tan, R.; Frankel, A. D.; Santi, D. V.; Cohen, F. E.; Bartlett, P. A. Peptoids: A Modular Approach to Drug Discovery. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9367–9371.

(23) Bartlett, P. A.; Santi, D. V.; Simon, R. J. Int. Patent WO91/19735, 1991.

(24) Zuckermann, R. N.; Kerr, J. M.; Kent, S. B. H.; Moos, W. H. Efficient method for the preparation of peptoids[oligo(N-substituted glycines)] by submonomer solid-phase synthesis. *J. Am. Chem. Soc.* **1992**, *114*, 10646–10647.

(25) Zuckermann, R. N. Presented at the Third International Symposium on Solid Phase Synthesis and Complementary Technologies, Oxford, U.K., 31 August–4 September, 1993.

(26) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. Encoded Combinatorial Peptide Libraries Containing Non-Natural Amino Acids. *J. Am. Chem. Soc.* **1993**, *115*, 2529–2531.

(27) Rich, D. H. Peptidase Inhibitors. In *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: Oxford, 1990, Vol. 2.

(28) Bartlett, P. A.; Marlowe, C. K. Possible Role for Water Dissociation in the Slow Binding of Phosphorous-Containing Transition-State-Analogue Inhibitors of Thermolysin. *Biochemistry* **1987**, *26*, 8553–8561.

(29) Morgan, B. P.; Scholtz, J. M.; Ballinger, M. D.; Zipkin, I. D.; Bartlett, P. A. Differential Binding Energy: A Detailed Evaluation of the Influence of Hydrogen-Bonding and Hydrophobic Groups on the Inhibition of Thermolysin by Phosphorous-Containing Inhibitors. *J. Am. Chem. Soc.* **1991**, *113*, 297–307.

(30) Campbell, D. A.; Bermak, J. C. Phosphonate Ester Synthesis Using a Modified Mitsunobu Condensation. *J. Org. Chem.* **1994**, *59*, 658–660. Campbell, D. A. The Synthesis of Phosphonate Esters, an Extension of the Mitsunobu Reaction. *J. Org. Chem.* **1992**, *57*, 6331–6335.

(31) Hagihara, M.; Anthony, N. J.; Stout, T. J.; Clardy, J.; Schreiber, S. L. Vinylogous Polypeptides: An Alternative Peptide Backbone. *J. Am. Chem. Soc.* **1992**, *114*, 6568–6570.

(32) Smith, A. B., III; Keenan, T. P.; Holcomb, R. C.; Sprengler, P. A.; Guzman, M. C.; Wood, J. L.; Carroll, P. J.; Hirschmann, R. Design, Synthesis and Crystal Structure of a Pyrrolinone-Based Peptidomimetic Possessing the Conformation of a β-Strand: Potential Applications to the Design of Novel Inhibitors of Proteolytic Enzymes. *J. Am. Chem. Soc.* **1992**, *114*, 10672–10674.

(33) Bunin, B. A.; Ellman, J. A. A General and Expedient Method for the Solid-Phase Synthesis of 1,4-Benzodiazepine Derivatives. *J. Am. Chem. Soc.* **1992**, *114*, 10997–10998.

(34) Ellman, J. A. Pharmaceutical Manufacturers Association, Drug Discovery Management Subsection, September 19–21, 1993, Philadelphia, PA.

(35) Nikolaiev, V.; Stierandova, A.; Krchnak, V.; Seligmann, B.; Lam, K. S.; Salmon, S. E.; Lebl, M. Peptide-Encoding for Structure Determination of Nonsequenceable Polymers Within Libraries Synthesized and Tested on Solid-Phase Supports. *Pept. Res.* **1993**, *6*, 161–170.

(36) Hobbs DeWitt, S.; Kiely, J. S.; Stankovic, C. J.; Schroeder, M. C.; Reynolds Cody, D. M.; Pavia, M. R. "Diversomers": An Approach to Nonpeptide, Nonoligomeric Chemical Diversity. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6909–6913.

(37) Mapelli, C.; Swerdloff, M. D. Monitoring of Conformational and Reaction Events in Resin-Bound Peptides by Carbon-13 NMR Spectroscopy in Various Solvents. In *Peptides. Proceedings of the 21st European Peptide Symposium*; Giralt, E. and Andreu, D., Eds., ESCOM, Leiden, The Netherlands, 1990; pp 316–319.

(38) Giralt, E.; Rizo, J.; Pedroso, E. Application of Gel-Phase ^{13}C-NMR to Monitor Solid Phase Peptide Synthesis. *Tetrahedron* **1984**, *40*, 4141–4152.

(39) Nicolaou, K. C.; Salvino, J. M.; Raynor, K.; Pietranico, S.; Reisine, T.; Freidinger, R. M.; Hirschmann, R. Design and Synthesis of a Peptidomimetic Employing β-D-Glucose for Scaffolding. *Peptides* **1989**, 881–884.

(40) Hirschmann, R.; Nicolaou, K. C.; Pietranico, S.; Salvino, J.; Leahy, E. M.; Sprengeler, P. A.; Furst, G.; Smith, A. B., III; Strader, C. D.; Cascieri, M. A.; Candelore, M. R.; Donaldson, C.; Vale, W.; Maechler, L. Nonpeptidal Peptidomimetics with a β-D-Glucose Scaffolding. A Partial Somatostatin Agonist Bearing a Close Structural Relationship to a Potent, Selective Substance P Antagonist. *J. Am. Chem. Soc.* **1992**, *114*, 9217–9218.

(41) Hirschmann, R.; Sprengeler, P. A.; Kawasaki, T.; Leahy, J. W.; Shakespeare, W. C.; Smith, A. B., III, The First Design and Synthesis of a Steroidal Peptidomimetic. The Potential Value of Peptidomimetics in Elucidating the Bioactive Conformation of Peptide Ligands. *J. Am. Chem. Soc.* **1992**, *114*, 9699–9701.

(42) Hirschmann, R.; Sprengeler, P. A.; Kawasaki, T.; Leahy, J. W.; Shakespeare, W. C.; Smith, A. B., III, The Versatile Steroid Nucleus: Design and Synthesis of a Peptidomimetic Employing this Novel Scaffold. *Tetrahedron* **1993**, *49*, 3665–3676.

(43) Parmley, S. F.; Smith, G. P. Antibody-Selectable Filamentous Fd Phage Vectors: Affinity Purification of Target Genes. *Gene* **1988**, *73*, 305–318.

(44) Barrett, R. W.; Cwirla, S. E.; Ackerman, M. S.; Olson, A. M.; Peters, E. A.; Dower, W. J. Selective Enrichment and Characterization of High Affinity Ligands from Collections of Random Peptides on Filamentous Phage. *Anal. Biochem.* **1992**, *204*, 357–364.

(45) Bass, S.; Green, R.; Wells, J. A. Hormone Phage: An Enrichment Method for Variant Proteins with Altered Binding Properties. *Proteins: Struct. Funct. Genet.* **1990**, *8*, 309–314.

(46) Devlin, J. J.; Panganiban, L. C.; Devlin, P. E. Random Peptide Libraries: A Source of Specific Protein Binding Molecules. *Science* **1990**, *249*, 404–406.

(47) Kay, B. K.; Adey, N. B.; He, Y.-S.; Manfredi, J. P.; Mataragnon, A. H.; Fowlkes, D. M. An M13 Phage Library Displaying Random 38-Amino-Acid Peptides as a Source of Novel Sequences with Affinity to Selected Targets. *Gene* **1993**, *128*, 59–65.

(48) Barrett, R. W.; Dower, W. J. Unpublished.

(49) Schatz, P. J. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: A 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Bio/Tech.* **1993**, *11*, 1138–1143.

(50) Doyle, M. Chiron Corporation. Presented at San Francisco Bay Area Peptide Society Meeting, December 1992.

(51) Lowman, H. B.; Bass, S. H.; Simpson, N.; Wells, J. A. Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* **1991**, *30*, 10832–10838.

(52) Garrard, L. J.; Yang, M.; O'Connell, M. P.; Kelley, R. F.; Henner, D. J. Fab Assembly and Enrichment in a Monovalent Phage Display System. *Bio/Technology* **1991**, *9*, 1373–1377.

(53) Hawkins, R. E.; Russell, S. J.; Winter, G. Selection of Phage Antibodies by Binding Affinity: Mimicking Affinity Maturation. *J. Mol. Biol.* **1992**, *226*, 889–896.

(54) Kang, A. S.; Barbas, C. F.; Janda, K. K.; Benkovic, S. J.; Lerner, R. A. Linkage of Recognition and Replication Functions by Assembling Combinatorial Antibody Fab Libraries Along Phage Surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 4363–4366.

(55) Christian, R. B.; Zuckermann, R. N.; Kerr, J. M.; Wang, L.; Malcolm, B. A. Simplified Methods for Construction, Assessment and Rapid Screening of Peptide Libraries in Bacteriophage. *J. Mol. Biol.* **1992**, *227*, 711–718.

(56) Zuckermann, R. N.; Kerr, J. M.; Siani, M. A.; Banville, S. C.; Santi, D. V. Identification of Highest-Affinity Ligands by Affinity Selection from Equimolar Peptide Mixtures Generated by Robotic Synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4505–4509.

(57) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. Identification of Antibody Mimotopes Containing Non-Natural Amino Acids by Recombinant and Synthetic Peptide Library Affinity Selection Methods. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 463–468.

(58) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **1990**, *249*, 505–510.

(59) Ellington, A. D.; Szostak, J. W. *In vitro* Selection of RNA molecules That Bind Specific Ligands. *Nature* **1990**, *346*, 818–822.

(60) Sherman, M. I.; Bertelsen, A. H.; Cook, A. F. Protein Epitope Targeting: Oligonucleotide Diversity and Drug Discovery. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 469–475.

(61) Irvine, D.; Tuerk, C.; Gold, L. Selexion: Systematic Evolution of Ligands by Exponential Enrichment with Integrated Optimization by Non-Linear Analysis. *J. Mol. Biol.* **1991**, *222*, 739–761.

(62) Tuerk, C.; MacDougal, S.; Gold, L. RNA Pseudoknots that Inhibit Human Immunodeficiency Virus Type 1 Reverse Transcriptase. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6988–6992.

(63) Brenner, S.; Lerner, R. A. Encoded Combinatorial Chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5181–5183.

(64) Nielsen, J.; Brenner, S.; Janda, K. D. Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **1993**, *115*, 9812–9813.

(65) Arnold, F. H.; Schofield, S. A.; Blanch, H. W. Analytical Affinity Chromatography: I. Local Equilibrium Theory and the Measurement of Association and Inhibition Constants. *J. Chromatogr.* **1986**, *355*, 1–12.

(66) Arnold, F. H.; Blanch, H. W. Analytical Affinity Chromatography: II. Rate Theory and the Measurement of Biological Binding Kinetics. *J. Chromatogr.* **1986**, *355*, 13–27.

(67) Zopf, D.; Ohlson, S. Weak-Affinity Chromatography. *Nature* **1990**, *346*, 87–88.

(68) Holmes, C. P. Polymer Reversal on Solid Surfaces. U.S. Pat. 5,242,974.

(69) Li, B.-L.; Langer, J. A.; Schwartz, B.; Petska, S. Creation of Phosphorylation Sites in Proteins: Construction of a Phosphorylatable Human Interferon α. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 558–562.

(70) Hollenbaugh, D.; Chalupny, N. J.; Aruffo, A. Recombinant Globulins: Novel Research Tools and Possible Pharmaceuticals. *Curr. Opinion Immunol.* **1992**, *4*, 216–219.

(71) Erickson, J.; Goldstein, B.; Holowka, D.; Baird, B. The Effect of Receptor Density on the Forward Rate Constant for Binding of Ligands to Cell Surface Receptors. *Biophys. J.* **1987**, *52*, 657–662.

(72) Jayawickreme, C. K.; Graminski, G. F.; Quillan, J. M.; Lerner, M. R. Creation and Functional Screening of a Multi-Use Peptide Library. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1614–1618.

(73) Ecker, D. J.; Vickers, T. A.; Hanecak, R.; Driver, V.; Anderson, K. Rational Screening of Oligonucleotide Combinatorial Libraries for Drug Discovery. *Nucleic Acids Res.* **1993**, *21*, 1853–1856.
(74) Blake, J.; Litzi-Davis, L. Evaluation of Peptide Libraries: An Iterative Strategy to Analyze the Reactivity of Peptide Mixture with Antibodies. *Bioconj. Chem.* **1992**, *3*, 510–513.
(75) Dooley, C. T.; Chung, N. N.; Schiller, P. W.; Houghten, R. A. Acetalins: Opioid Receptor Antagonists Determined Through the Use of Synthetic Peptide Combinatorial Libraries. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10811–10815.
(76) Houghten, R. A.; Appel, J. R.; Blondelle, S. E.; Cuervo, J. H.; Dooley, C. T.; Pinilla, C. The Use of Synthetic Peptide Combinatorial Libraries for the Identification of Bioactive Peptides. *BioTechniques* **1992**, *13*, 412–421.
(77) Salmon, S. E.; Lam, K. S.; Lebl, M.; Kandola, A.; Khattri, P. S.; Wade, S.; Patek, M.; Kocis, P.; Krchnak, V.; Thorpe, D.; Felder, S. Discovery of Biologically Active Peptides in Random Libraries: Solution-Phase Testing After Staged Orthogonal Release from Resin Beads. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11708–11712.
(78) Gordon, E. M.; Godfrey, J. D.; Pluscec, J.; Von Langen, D.; Natarajan, S. Design of Peptide Derived Amino Alcohols as Transition-State Analog Inhibitors of Angiotensin Converting Enzyme. *Biochem. Biophys. Res. Commun.* **1985**, *126*, 419–426.
(79) Godfrey, J. D.; Gordon, E. M.; Von Langen, D.; Engebrecht, J. Synthesis of Peptidyl Derived Amino Alcohols. *J. Org. Chem.* **1986**, *51*, 3073.
(80) Rich, D. H.; Green, J.; Toth, M. V.; Marshall, G. R.; Kent, S. B. H. Hydroxyethylamine Analogs of the p17/p24 Substrate Cleavage Site are Tight-Binding Inhibitors of HIV Protease. *J. Med. Chem.* **1990**, *33*, 1285–1288.
(81) Affymetrix Co. of Santa Clara, CA, a subsidiary of Affymax, uses the VLSIPS process for creation of oligonucleotide libraries on chips for the purposes of genetic typing and high-through-put DNA sequencing.

Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 6027–6031, June 1995
Pharmacology

EXHIBIT 3

A paradigm for drug discovery employing encoded combinatorial libraries

(medicinal chemistry/isozyme selectivity/combinatorial chemistry)

JONATHAN J. BURBAUM*, MICHAEL H. J. OHLMEYER, JOHN C. READER, IAN HENDERSON, LAWRENCE W. DILLARD, GE LI, TROY L. RANDLE, NOLAN H. SIGAL, DANIEL CHELSKY, and JOHN J. BALDWIN

Departments of Biology and Chemistry, Pharmacopeia, Inc., 101 College Road East, Princeton, NJ 08540

Communicated by Michael H. Wigler, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, February 8, 1995 (received for review November 1, 1994)

ABSTRACT **Very large combinatorial libraries of small molecules on solid supports can now be synthesized and each library element can be identified after synthesis by using chemical tags. These tag-encoded libraries are potentially useful in drug discovery, and, to test this utility directly, we have targeted carbonic anhydrase (carbonate dehydratase; carbonate hydro-lyase, EC 4.2.1.1) as a model. Two libraries consisting of a total of 7870 members were synthesized, and structure–activity relationships based on the structures predicted by the tags were derived. Subsequently, an active representative of each library was resynthesized {2-[*N*-(4-sulfamoylbenzoyl)-4′-aminocyclohexanespiro]-4-oxo-7-hydroxy-2,3-dihydrobenzopyran and [*N*-(4-sulfamoylbenzoyl)-L-leucyl]piperidine-3-carboxylic acid} and these compounds were shown to have nanomolar dissociation constants (15 and 4 nM, respectively). In addition, a focused sublibrary of 217 sulfamoylbenzamides was synthesized and revealed a clear, testable structure–activity relationship describing isozyme-selective carbonic anhydrase inhibitors.**

The current national focus on health care reform has highlighted the continuing importance of developing more cost-effective methods for treating disease. Historically, the pharmaceutical industry has contributed to this effort by discovering and developing therapeutic molecules. The development of drug candidates is expensive (1, 2), and preclinical research toward the discovery and development of lead structures contributes significantly to this expense. For example, in 1991, the costs of medicinal chemistry and biological testing prior to safety assessment amounted to nearly 30% of domestic research costs (3). Apart from the economics of drug discovery, classical medicinal chemistry cannot efficiently address the plethora of new biochemical targets suggested by recent discoveries in molecular genetics, and lead structures from natural sources are often too complex for the cost-effective synthesis of analogs. Hence, targets for therapeutic intervention cannot be rapidly and efficiently exploited.

Combinatorial chemical synthesis applied to drug discovery promises to improve the productivity of medicinal chemistry both by significantly increasing the number of molecules available for testing and by providing facile routes toward synthetic analogs of active molecules. Many strategies for the generation of chemical diversity have been proposed (4–7). With a single exception (4), these approaches have been confined to the synthesis of flexible oligomeric ligands of relatively high molecular weight (e.g., peptides and oligonucleotides) which tend to be poor therapeutic agents, in part because of their lack of availability and stability *in vivo*.

For drug discovery, small-molecule libraries containing diverse functionality in a variety of molecular scaffolds possess the greatest utility. To construct such structurally diverse libraries while allowing the assignment of a structure to each member, we use a set of chemically stable molecular tags during solid-phase synthesis on polymeric beads (8). These tags can be used to unambiguously encode each bead with the synthetic scheme for its library member and are sufficiently robust to allow a wide range of chemical reactions during ligand construction.

A drug discovery effort based on such combinatorial libraries must pass several tests to be successful. First, the library must be diverse enough to permit the identification of a small subset of active molecules among many less active or inactive ones. Second, the most promising candidates in the library should suggest a testable structure–activity relationship. Third, since selectivity for a particular target is often important, strategies for the analysis of libraries containing a significant fraction of active compounds should result in optimization of selectivity as well as activity.

To explore the capabilities of encoded combinatorial libraries in a small-molecule drug discovery effort directed toward lead identification and optimization, we chose carbonic anhydrase (carbonate dehydratase; carbonate hydro-lyase, EC 4.2.1.1) as a therapeutically significant model, since inhibitors of carbonic anhydrase are known to be useful in ameliorating the symptoms of glaucoma. The inhibition of carbonic anhydrase by compounds containing primary sulfamoyl groups ($-SO_2NH_2$) has been well-characterized, both pharmacologically (9–11) and structurally (12–15). This allowed the design of test libraries that were predisposed toward showing carbonic anhydrase inhibition without explicitly including known inhibitors. Finally, since a number of different forms of the enzyme have been characterized, we were able to perform experiments to probe isoform selectivity.

MATERIALS AND METHODS

Materials. TentaGel S-NH_2 resin (particle size, 130 μm) was obtained from Rapp Polymere (Tubingen, Germany). The photocleavable linker was synthesized as described (16). Carbonic anhydrase was obtained from Sigma. White 96-well polystyrene plates were obtained from Dynatech.

Library Synthesis and Analysis. Libraries were synthesized by the split synthesis protocol, and each bead thus contains one member of the library (17–19). All members were affixed via a photocleavable *o*-nitrobenzyl linker. To preserve the synthetic history of each bead, each library was indexed in a binary fashion using electrophoric tags. These tags were attached by using $[(CF_3CO_2)_2Rh]_2$-activated carbene insertion as described (20). After identification and isolation of beads, the tags were analyzed by gas chromatography (8).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: DNSA, dansylamide; hCA(I) and hCA(II), human carbonic anhydrase isozymes I and II.

*To whom reprint requests should be addressed.

Synthesis of Dihydrobenzopyran Library. First, three dihydroxyacetophenones (2-acetyl-1,4-dihydroxybenzene, 2-acetyl-1,5-dihydroxybenzene, and 2-acetyl-6-methyl-1,5-dihydroxybenzene) were coupled using Mitsunobu conditions (triphenylphosphine/diethylazodicarboxylate in tetrahydrofuran at ambient temperature) to the *o*-nitrobenzyl linker, which was then coupled to the beads by carbodiimide coupling [di(isopropylcarbodiimide)/hydroxybenzotriazole in dimethylformamide]. These intermediates were cyclized on the resin with a set of seven ketones, four of which contained a protected amine functionality. Modification of the amine functionality (when present) with 31 different headgroups under standard conditions produced a library of ketones containing 381 members. The ketones were then reduced with $NaBH_4$, converted into the corresponding dithiolanes with ethane-1,2-dithiol, or left unaltered to produce a library of 1143 distinct compounds.

Synthesis of Acylpiperidine Library. Library elements were coupled as either a carboxylic ester or a carbonate ester of an *o*-nitrobenzyl alcohol. First, five piperidines (2-hydroxymethyl-, 3-hydroxymethyl-, 4-hydroxy-, 4-carboxylic-, and 3-carboxylic-), and two linear moieties (5-amino-3-oxopentanol and 6-aminohexanoic acid) were attached to this linker either by phosgene coupling in solution (for the alcohols) or directly to the resin (for the acids). Next, 31 amino acids, including 28 examples of both D and L isomers of the 20 commonly occurring α-amino acids, and 3 nonproteinogenic amino acids (piperidine-4-carboxylic acid, piperidine-3-carboxylic acid, and 6-aminohexanoic acid) were reacted with the free amine. Finally, 31 reagents were linked as carboxamides, carbamates, sulfonamides, and ureas, including three primary sulfonamides (4-sulfamoylbenzamido, 4-chloro-5-sulfamoylbenzamido, and 2,4-dichloro-5-sulfamoylbenzamido).

Library Analysis for Carbonic Anhydrase Inhibition. Beads were distributed into 96-well microtiter plates. Compounds were detached from the beads by UV irradiation (350 nm) and then transferred to assay plates. An assay solution (50 μl) consisting of 0.1 M phosphate buffer (pH 7.4) containing bovine carbonic anhydrase (0.3 μM) and dansylamide (DNSA; 0.6 μM) was added to each well. Fluorescence values (λ_{ex}, 274 nm; λ_{em}, 454 nm) were measured with a Perkin–Elmer model LS 50B spectrofluorimeter equipped with a microtiter plate reader accessory and were normalized (uninhibited, 100; empty, 0). Assay solutions from the initial screen identified as active were transferred to a second fluorescence plate and 5 μl of 2 mM DNSA in dimethyl sulfoxide was added to each well to increase [DNSA] ≈600-fold *in situ*. Under these conditions, chlorothiazide ($K_d \approx 75$ nM) was displaced, while acetazolamide ($K_d \approx 7.5$ nM) was not, suggesting that more potent inhibitors were less sensitive to displacement by elevated [DNSA].

Selectivity Analysis. Aliquots (one-third) from the same bead eluate were assayed against each isozyme. Concentrations of the two isozymes were matched using $\varepsilon_{280}^{1\%}$ values from the literature (21), and the [DNSA] was varied according to the literature values for K_d (DNSA) of the isozymes (22).

RESULTS AND DISCUSSION

Two model libraries for drug discovery—a 1143-member dihydrobenzopyran library and a 6727-member acylpiperidine library—were designed. The first (dihydrobenzopyran) library (Fig. 1*A*) was designed to include 4-sulfamoylbenzoyl amides that were expected to confer carbonic anhydrase activity. The second (acylpiperidine) library (Fig. 1*B*) was designed around a core aminocarboxamide derived from 31 amino acids, including nonproteinogenic amino acids and both the D and L configurations of several naturally occurring α-amino acids. The carboxamide was formed from the carboxylic acid with seven amines including four substituted piperidines. Three different sulfamoylbenzoyl groups were incorporated at the final step to confer carbonic anhydrase activity.

Activity of library members against bovine carbonic anhydrase was assessed by using a fluorescence-based ligand-displacement assay (9). Because each synthetic library contained >1000 beads per member, we characterized the activity of the library statistically to optimize our screening conditions. The dihydrobenzopyran library was determined to contain ≈1.4% actives, identified as those wells where the fluorescence decreased by $>5\sigma$. Similarly, in the acylpiperidine library, we determined that the samples contained ≈6% actives. These observations were consistent with the percentage of primary sulfonamides incorporated in the two libraries, which was predicted to be 3% and 10% for the dihydrobenzopyran and acylpiperidine libraries, respectively.

Over 2300 beads from the dihydrobenzopyran library were then assayed singly and 33 individual beads were chosen for decoding. Thus, 2.0 library equivalents were assayed, where one library equivalent is defined as the number of beads equal to the number of distinct library members. The probability that a given compound was not assayed is approximated as follows: In a library of N total beads $\gg M$, the total number of library members, when S beads have been assayed, the probability that a given member has not been chosen is given by

$$P_s(0) \approx \left[1 - \frac{1}{M}\right]^S, \qquad [1]$$

which is approximately $(e^{-1})^{S/M}$ for large S. The quantity S/M is L, the number of library equivalents that have been assayed, so $P_s(0) \approx e^{-L}$. For the case at hand, the fraction of the library assayed at least once is $(1 - e^{-2})$ or 86%.

The larger acylpiperidine library was pooled at the final reaction step into two smaller portions to permit a more convergent assay strategy. These pools consisted of 3472 and 3255 members, the larger portion containing all the primary sulfonamides. For the smaller portion, ≈5 library equivalents (≈17,000 beads) were assayed at 10 beads per well, and no actives were identified. Thus, in Eq. **1**, $L \approx 5$ and $P_s(0)$ is <0.01. Hence, >99% of the members were assayed at least once, and none were found to be active under the conditions described.

We therefore focused our continued efforts on the portion of the acylpiperidine library that contained primary sulfonamides. Here, half the eluates from a total of 4320 single beads (1.3 library equivalents) were assayed individually, and >300 actives were identified (Fig. 2*A*). To estimate the relative potency of the active inhibitors, the stringency of the assay was increased by raising the concentration of the competitor, DNSA. However, the associated background fluorescence made the fluorescence decrease caused by inhibitors less easily detected. Potent actives were therefore judged as differing by $>3\sigma$ compared to the median active compound. Eighteen high-affinity actives were thereby identified and decoded (Fig. 2*B*).

As anticipated, we found that carbonic anhydrase inhibitors from either library exclusively contained the sulfamoyl group ($-SO_2NH_2$). From the actives, we chose two synthetic targets, one from the dihydrobenzopyran library (compound **1**) and one of the more potent actives from the acylpiperidine library (compound **2**). These compounds were resynthesized, their structures were confirmed spectroscopically, and their K_d values vs. carbonic anhydrase were determined experimentally (Table 1). The low nanomolar values for compounds **1** and **2** support the use of tag-derived information for structure–activity assessments in the libraries we report.

Significantly, no structures containing a 4-chloro-5-sulfamoylbenzoyl moiety were found among the tag-predicted structures for potent actives in the acylpiperidine library. Furthermore, while the other two sulfamoylbenzamides were

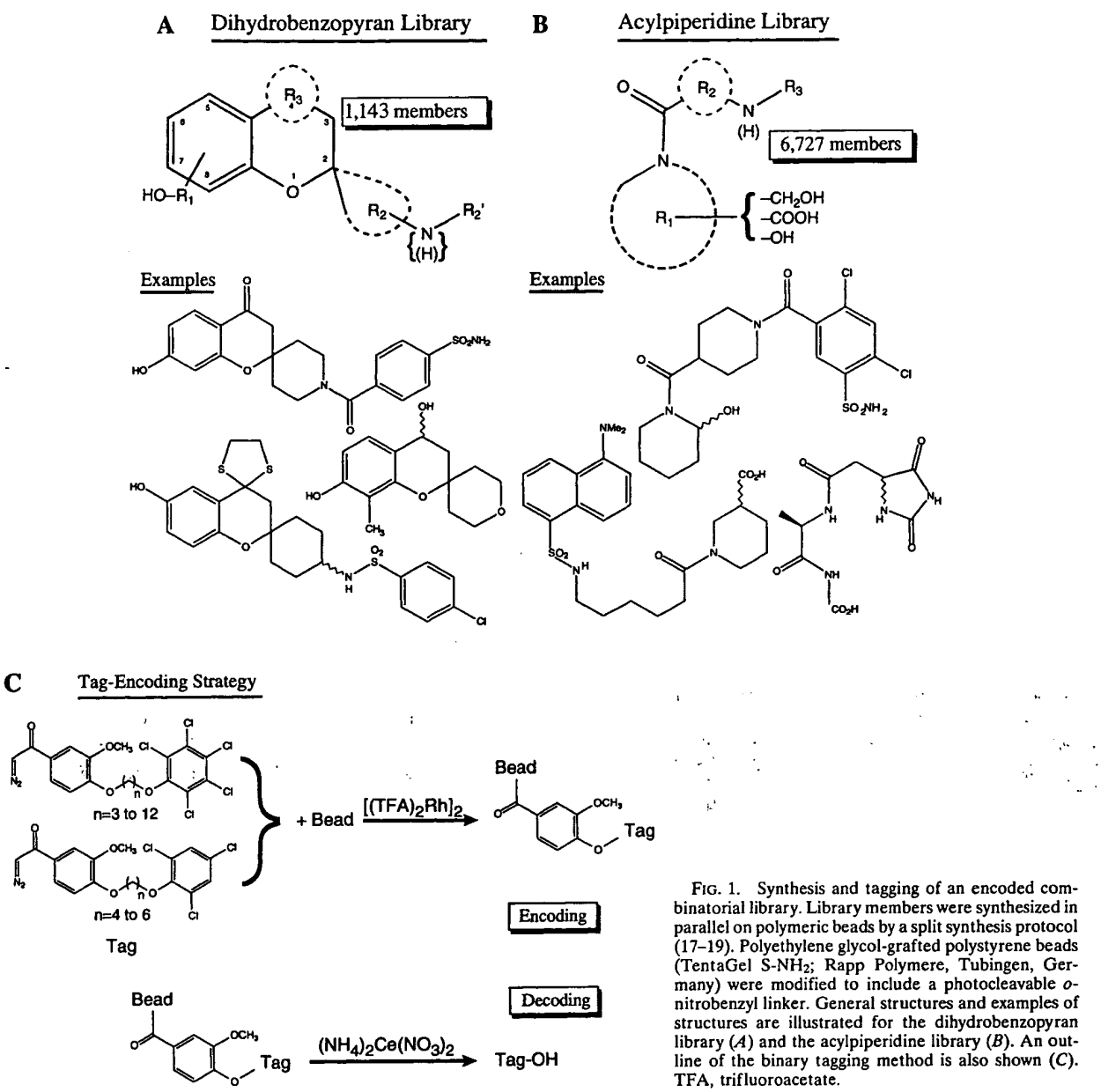


FIG. 1. Synthesis and tagging of an encoded combinatorial library. Library members were synthesized in parallel on polymeric beads by a split synthesis protocol (17–19). Polyethylene glycol-grafted polystyrene beads (TentaGel S-NH_2; Rapp Polymere, Tubingen, Germany) were modified to include a photocleavable *o*-nitrobenzyl linker. General structures and examples of structures are illustrated for the dihydrobenzopyran library (*A*) and the acylpiperidine library (*B*). An outline of the binary tagging method is also shown (*C*). TFA, trifluoroacetate.

well-represented, they showed markedly different structure–activity relationships. For example, the 4-sulfamoylbenzamides appear to prefer a hydrophobic R_2 group of the L configuration, while the 2,4-dichloro-5-sulfamoylbenzamides appear to tolerate more diversity at R_2, including D-alanine. Thus, to test the validity of this observation, library-specific homologs of the acylpiperidine (compound **2**) were synthesized (Table 1, compounds **3** and **4**). The 150-fold increase in K_d seen upon substitution further validates the ability of the screen to select active compounds and supports the divergent structure–activity relationships observed with the various sulfonamide headgroups.

In developing a class of compounds for use as therapeutic agents, increased affinity is but one criterion for optimization. Isozyme selectivity is often an additional critical property required to minimize side effects. The selectivity of a large library is difficult to optimize convergently (i.e., with >1 element per assay), because such assays must measure the absence of activity against a particular target or relative levels of activity vs. multiple targets. As the percentage of actives increases, the probability that the observed affinities are the property of multiple elements increases. Thus, approaches taken to identify high-affinity peptides in a convergent fashion (e.g., see ref. 6) become impractical with selectivity, and focused sublibraries assayed singly become pertinent. To evaluate the potential for using encoded combinatorial libraries to optimize selectivity, a 217-member sublibrary (consisting of the 4-sulfamoylbenzamides from the larger acylpiperidine library) was prepared and evaluated against human carbonic anhydrase isozymes I and II [CA(I) and CA(II)].

To isolate selectivity as a variable, two factors were normalized. First, equal inhibitor concentrations were exposed to each isozyme by taking aliquots from the same bead eluate. Thus, one-third of a complete eluate was assayed vs. human CA(I), and one-third was assayed vs. CA(II). Second, the concentrations of the two isozymes were matched, but the

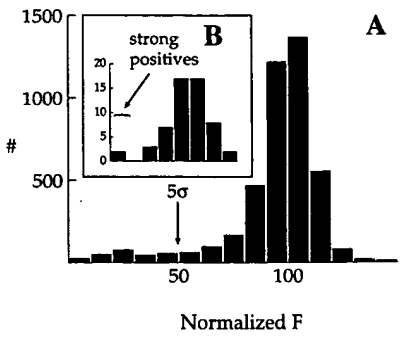


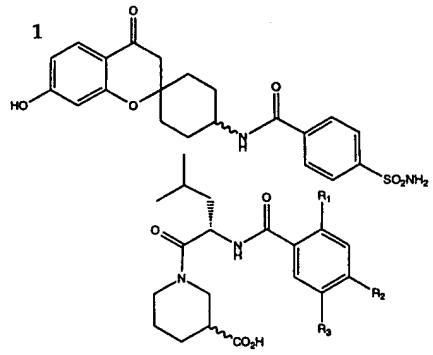
FIG. 2. · Histogram analysis of combinatorial libraries. (*A*) Low-stringency library analysis. Half the eluates from a total of 4320 single beads (1.2 library equivalents) were assayed individually, and >300 actives (5σ) were identified. For the dihydrobenzopyran library, >2300 members (2.0 library equivalents) were assayed singly and 33 individual beads were identified and decoded. (*B*) Reanalysis of positives from *A* at higher stringency. To increase the stringency of the assay, the concentration of DNSA was increased ≈600-fold *in situ*, and 18 potent actives were selected for decoding.

[DNSA] was varied according to the literature values for K_{DNSA} of the isozymes. These requirements derive from the binding equation, given relative to isozyme x

$$F_x \propto [E_x\text{·DNSA}] = \frac{[E]_{tot,x}}{1 + \frac{K_{DNSA,x}}{[DNSA]_x}\left(1 + \frac{[I]}{K_{i,x}}\right)}. \quad [2]$$

Thus, when E_{tot} is equal, the ratio $[DNSA]/K_{DNSA}$ is equal, and [I] is equal, the only variable that can affect F is the dissociation constant for the inhibitor, K_i. To identify samples for which there is a significant difference in K_i for the two isozymes, the normalized fluorescence values were analyzed graphically (Fig. 3).

Table 1. Analysis of compounds and validation of tag-derived structure–activity relationships

Compound				K_d[bCA(II)], nM
1				15
	R_1	R_2	R_3	
2	H	SO_2NH_2	H	4
3	H	Cl	SO_2NH_2	700
4	Cl	Cl	SO_2NH_2	660

Structures of compounds **1–4** were confirmed analytically by ^1H NMR spectroscopy and mass spectrometry. K_d values were measured as described (7) with EXCEL software to solve the multiple equilibria numerically. b, Bovine.

Fig. 3*A* shows a selectivity plot comparing the normalized fluorescence values for the two enzymes. In this figure, region A contains compounds that show potency vs. both isozymes, while region B contains compounds that are less selective for CA(II) than the median, and region C contains compounds that are more selective than the median. Sulfonamides that are not particularly active against either isozyme are found in region D. Beads from each region were decoded, and examples of the structures predicted by these codes are shown in Fig. 2*B*. The potent compound that was identified in screening against bovine carbonic anhydrase (Table 1, compound **2**) was iden-

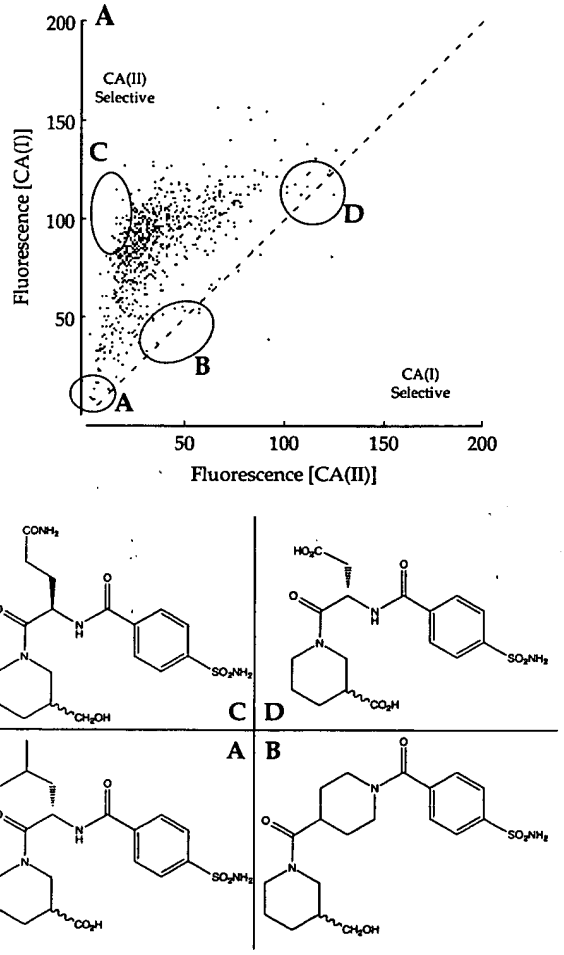


FIG. 3. Selectivity assay of a highly active sublibrary. (*A*) Inhibitors were analyzed independently for each isozyme as in Fig. 2 and then plotted. The observation that the majority of the library members are found above the diagonal suggests that the selectivity for hCA(II) may be a property of this class of compounds rather than a characteristic of a given member. (*B*) From tag-derived relationships, there are clear trends associated with the different regions, and some examples of compounds are shown. The characteristics of the potent inhibitors (region A) are described above, and the potent compound that was identified in assays against bovine carbonic anhydrase (Table 1, compound 2) was found here. Region B (relatively nonselective) shows a preference for cyclic imino acids in the central region, which may reflect differential hydrogen bonding by the carboxamido nitrogen of the sulfamoylbenzamide (13) or to a positive discrimination by the ring. Region C [relatively hCA(II) selective] shows a preponderance of polar side chains, as well as a relative lack of sensitivity toward the C^{α} configuration. Region D (low activity) contains more polar compounds but, notably, one D-leucyl side chain that complements the strong preference for hydrophobic groups of the L stereoconfiguration noted above.

tified again in region A. Analysis of the selectivity of this compound showed that it has a 3-fold selectivity for human (h) CA(II) {K_d[hCA(II)] = 11 nM; K_d[hCA(I)] = 33 nM}, but this fact would not be apparent from the plot because of its relatively high affinity. Since the other compounds in this small sublibrary are also analogs of 4-sulfamoylbenzamide, the observation that most library members are found above the diagonal suggests that the selectivity observed for compound **2** may be a property of this class of compounds rather than a characteristic of this particular member. Nevertheless, there are clear trends associated with the different regions. Region B (relatively nonselective) shows a preference for cyclic imino acids in the central region, which may reflect differential hydrogen bonding by the carboxamido nitrogen of the sulfamoylbenzamide (13), or to a positive discrimination by the ring. Region C [relatively hCA(II) selective] shows a preponderance of polar side chains as well as a relative lack of sensitivity toward the C^{α} configuration. Region D (low activity) contains more polar compounds but, notably, one D-leucyl side chain that complements the strong preference for hydrophobic groups of the L stereoconfiguration noted above. Together, the trends derived from library analysis suggest that more selective compounds could be synthesized, either as single compounds or as appropriately designed focused libraries.

Conclusions. We have synthesized encoded libraries containing 10^3–10^4 small molecules by using diverse chemistries. Active carbonic anhydrase inhibitors within these libraries have been identified and optimized by using directed sublibraries based on the original lead and the chemistry developed for its synthesis. This application of combinatorial libraries for drug discovery addresses two pressing needs of the pharmaceutical industry. First, small-molecule leads will become more readily available as the size and diversity of synthetic libraries increases. This, in turn, will accelerate the discovery of compounds active against a variety of therapeutic targets, especially those that emerge from the explosion in sequence information arising from the human genome project. Second, synthetic chemists involved in drug discovery will be more efficiently engaged to solve sophisticated structure–activity problems. The success of the combinatorial approach emphasizes the power of techniques for the rapid synthesis and evaluation of large numbers of possible structures and anticipates a new age of drug discovery and optimization. In the future, analysis of combinatorial libraries may be used to direct research efforts toward productive avenues by suggesting the synthesis either of a handful of pertinent test compounds or of combinatorial libraries based on relevant structural types. Indeed, such a combinatorial approach may be productively applied not just in pharmaceutical discovery, but in any area where optimization of molecular properties is necessary.

We wish to thank Pharmacopeia's engineering department, in particular Joseph Brzezinski and Jeffrey Davidson, for providing plates containing arrayed beads, and Peter Kieselbach and Nicole White for electron capture gas chromatography analysis of tagged beads. We also thank Laura Rokosz (Department of Biology, Pharmacopeia) for expert HPLC analysis. We thank Dr. Harvey Schwam, Professor W. Clark Still (Columbia University, New York), Professor Paul D. Bartlett (University of California, Berkeley), and Professor Tom Maren (University of Florida) for helpful scientific discussions and John Chabala, Gabriel Lopez, and Joseph Mollica (Pharmacopeia) for careful reading of the manuscript. Michael H. Wigler is a consultant to Pharmacopeia in the area of combinatorial chemical synthesis.

1. Vagelos, P. R. (1991) *Science* **252,** 1080–1084.
2. Kirschner, M. W., Marincola, E. & Teisberg, E. O. (1994) *Science* **266,** 49–51.
3. Pharmaceutical Research and Manufacturers of America (1993) *PMA Annual Survey Report: Trends in U. S. Pharmaceutical Sales and R & D* (Pharmaceut. Manuf. Am., Washington, DC), p. 29.
4. Bunin, B. A. & Ellman, J. A. (1992) *J. Am. Chem. Soc.* **114,** 10997–10998.
5. Needels, M. C., Jones, D. G., Tate, E. H., Heinkel, G. L., Kochersperger, L. M., Dower, W. J., Barrett, R. W. & Gallop, M. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10700–10704.
6. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T. & Cuervo, J. H. (1991) *Nature (London)* **354,** 84–86.
7. Simon, R. J., Kania, R. S., Zuckermann, R. N., Huebner, V. D., Jewell, D. A., Banville, S., Ng, S., Wang, L., Rosenberg, S., Marlowe, C. K., Spellmeyer, D. C., Tan, R., Frankel, A. D., Santi, D. V., Cohen, F. E. & Bartlett, P. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9367–9371.
8. Ohlmeyer, M. H. J., Swanson, R. N., Dillard, L. W., Reader, J. C., Asouline, G., Kobayashi, R., Wigler, M. & Still, W. C. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10922–10926.
9. Ponticello, G. S., Freedman, M. B., Habecker, C. N., Lyle, P. A., Schwam, H., Varga, S. L., Christy, M. E., Randall, W. C. & Baldwin, J. J. (1987) *J. Med. Chem.* **30,** 591–597.
10. Hartman, G. D., Halczenko, W., Smith, R. L., Sugrue, M. F., Mallorga, P. J., Michelson, S. R., Randall, W. C., Schwam, H. & Sondey, J. M. (1992) *J. Med. Chem.* **35,** 3822–3831.
11. Lo, Y. S., Nolan, J. C., Maren, T. H., Welstead, W. J., Jr., Gripshover, D. F. & Shamblee, D. A. (1992) *J. Med. Chem.* **35,** 4790–4794.
12. Jain, A., Whitesides, G. M., Alexander, R. S. & Christianson, D. W. (1994) *J. Med. Chem.* **37,** 2100–2105.
13. Cappalonga Bunn, A. M., Alexander, R. S. & Christianson, D. W. (1994) *J. Am. Chem. Soc.* **116,** 5063–5068.
14. Ericksson, A. E., Kylsten, P. M., Jones, T. A. & Liljas, A. (1988) *Proteins Struct. Funct. Genet.* **4,** 283–293.
15. Smith, G. M., Alexander, R. S., Christianson, D. W., McKeever, B. M., Ponticello, G. S., Springer, J. P., Randall, W. C., Baldwin, J. J. & Habecker, C. N. (1994) *Protein Sci.* **3,** 118–125.
16. Barany, G. & Albericio, F. (1985) *J. Am. Chem. Soc.* **107,** 4936–4942.
17. Furka, Á., Sebesteyén, F., Asgedom, M. & Dibó, G. (1991) *Int. J. Peptide Protein Res.* **37,** 487–493.
18. Sebesteyén, F., Dibó, G., Kovacs, A. & Furka, Á. (1993) *Bioorg. Med. Chem. Lett.* **3,** 413–418.
19. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M. & Knapp, R. J. (1991) *Nature (London)* **354,** 82–84.
20. Nestler, H. P., Bartlett, P. A. & Still, W. C. (1994) *J. Org. Chem.* **59,** 4723–4724.
21. Nyman, P. O. & Lindskog, S. (1964) *Biochim. Biophys. Acta* **85,** 141–151.
22. Taylor, P. W., King, R. W. & Burgen, R. S. V. (1970) *Biochemistry* **9,** 2638–2645.

Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 5072–5076, May 1995
Biochemistry

EXHIBIT 4

Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE

KARL R. CLAUSER*, STEVEN C. HALL*, DIANA M. SMITH†, JAMES W. WEBB*‡, LORI E. ANDREWS*, HUU M. TRAN†, LOIS B. EPSTEIN†§¶, AND ALMA L. BURLINGAME*‖**

*Department of Pharmaceutical Chemistry, †Cancer Research Institute, §Department of Pediatrics, and ¶Liver Center, University of California, San Francisco, CA 94143

Communicated by John A. Clements, University of California, San Francisco, CA, January 18, 1995 (received for review June 29, 1994)

ABSTRACT We report a general mass spectrometric approach for the rapid identification and characterization of proteins isolated by preparative two-dimensional polyacrylamide gel electrophoresis. This method possesses the inherent power to detect and structurally characterize covalent modifications. Absolute sensitivities of matrix-assisted laser desorption ionization and high-energy collision-induced dissociation tandem mass spectrometry are exploited to determine the mass and sequence of subpicomole sample quantities of tryptic peptides. These data permit mass matching and sequence homology searching of computerized peptide mass and protein sequence data bases for known proteins and design of oligonucleotide probes for cloning unknown proteins. We have identified 11 proteins in lysates of human A375 melanoma cells, including: α-enolase, cytokeratin, stathmin, protein disulfide isomerase, tropomyosin, Cu/Zn superoxide dismutase, nucleoside diphosphate kinase A, galaptin, and triosephosphate isomerase. We have characterized several post-translational modifications and chemical modifications that may result from electrophoresis or subsequent sample processing steps. Detection of comigrating and covalently modified proteins illustrates the necessity of peptide sequencing and the advantages of tandem mass spectrometry to reliably and unambiguously establish the identity of each protein. This technology paves the way for studies of cell-type dependent gene expression and studies of large suites of cellular proteins with unprecedented speed and rigor to provide information complementary to the ongoing Human Genome Project.

Cloning of genes associated with malignancy generates inevitable excitement in biology and medicine. However, subsequent study at the protein level is clearly necessary to understand the processes by which those genes affect vital biological functions, such as the control of gene expression and regulation of cell-signaling pathways. Two-dimensional (2D) PAGE is preferred for simultaneous separation and visualization of proteins present in cell lysates (1, 2). Protein identification and characterization of functionally important primary structural features are the first steps toward gaining insight into the biological roles of specific proteins (3–6). Furthermore, characterization of the 2D-PAGE map from a cell system facilitates both basic research and clinical diagnosis (3, 7).

Traditional partial-sequencing approaches for identifying proteins isolated by 2D PAGE, such as Edman degradation of electroblotted proteins, often meet with limited success due to N-terminal blockage of many eukaryotic proteins. While peptide cleavage, extraction, and HPLC separation may enable subsequent Edman sequencing of internal peptides (8), these procedures alone are not rapid enough to identify the thousands of proteins in a cell lysate in a timely manner. Because of their speed, sensitivity, and ability to deal directly with mixtures, several recently developed mass spectrometry techniques are rapidly becoming the primary methods for identifying proteins isolated by 2D PAGE (4–6, 9, 10). Previously, our laboratories reported the advantages of using liquid secondary ion mass spectrometry, high-energy collision-induced dissociation (CID) tandem mass spectrometry, Edman sequencing, and 2D PAGE to characterize lipocortin I from human melanoma lysates (4). Peptide mass determination and CID sequencing using sample quantities of ≈100 pmol revealed an acetylated N terminus and an unanticipated acrylamide-modified cysteine.

Recently, we have substantially reduced the detection limits of peptide sequencing through incorporation of continuous flow sample introduction and a scanning charge-coupled device array detector onto our tandem mass spectrometer. The resulting chemical noise reduction and rapid recording of single CID spectra (≈10 s) now enable routine peptide sequencing from sample quantities of 100 fmol to 10 pmol (11, 12). Prior to CID, substantial sample is conserved by exploiting the 1–100 fmol sensitivity of matrix-assisted laser desorption ionization (MALDI), rather than liquid secondary ion mass spectrometry for preliminary mass measurement. MALDI enables the generation of peptide molecular-mass maps from fmol levels of unfractionated protein digests or individual HPLC fractions. These peptide-mass fingerprints may be used to search peptide-mass data bases and predict protein identities without resorting to sequencing (5, 9, 13). However, mass-matching algorithms can find only proteins which are already present in a data base. Obviously, primary structure elucidation is essential to identify unknown proteins, search sequence data bases for homologous proteins, characterize covalent modifications, or design oligonucleotide probes for gene cloning. Taking these points into consideration, we report an integrated strategy for identifying and characterizing primary structural features of proteins isolated by 2D PAGE.

MATERIALS AND METHODS

Protein Isolation, Purification, Digestion, and Edman Sequencing. Preparation of human A375 melanoma lysates, isolation of proteins by 2D PAGE (4, 14), protein electroelution of pooled gel spots, tryptic digestion, HPLC separation, and Edman sequencing were performed as described (4). The procedure of Rosenfeld *et al.* (15) was used to produce an in-gel tryptic digest of spot 42. Control digestions without protein substrate were performed so that trypsin autoproteolysis products could be disregarded.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: 2D, two-dimensional; CID, collision-induced dissociation; MALDI, matrix-assisted laser desorption ionization.

‡Present address: Department of Chemistry, Illinois State University, Normal, IL 61790.
¶To whom reprint requests should be sent at the † address.
**To whom reprint requests should be sent at the * address.

Mass Spectrometry. Molecular masses (isotopic average) of all tryptic peptides were determined by analyzing 1/50th of each HPLC fraction with a VG TofSpec MALDI mass spectrometer equipped with a nitrogen laser and operated in the linear mode. Peptides were crystallized in matrices consisting of 100 mM 2,4-dihydroxybenzoic acid/50 mM fucose or a saturated solution of α-cyano-4-hydroxycinnamic acid prepared in 0.1% aqueous trifluoroacetic acid (TFA). The mean mass from all spectra recorded for a particular peptide is reported. All MALDI spectra were externally calibrated by using a standard peptide mixture. High-energy positive ion CID mass spectra were acquired on a Kratos Analytical Instruments Concept IIHH four-sector tandem mass spectrometer equipped with a continuous flow, liquid inlet probe and a scanning charge-coupled device array detector (11). Both MS1 and MS2 were operated at 1000 resolution ($M/\Delta m$) to determine monoisotopic masses. HPLC fractions were concentrated to ≈5 μl and diluted to ≈15 μl with a mixture of aqueous 5% (vol/vol) thioglycerol/5% (vol/vol) acetonitrile/0.1% TFA matrix solution. Samples were introduced into the mass spectrometer source at a flow rate of 3 μl/min. Sequences were deduced from CID spectra by using interactive interpretation software developed in our laboratory (16). Quantitation estimates were based on CID and UV response of standards.

Immunoblot Analysis. 2D immunoblot analysis was performed as described (4) with modifications. Gels were loaded with 400–800 μg of protein lysate and transferred at 1.3 mA/cm² for 2 h, and blots were stained with 0.1% Ponceau S/5% (vol/vol) acetic acid. Stain was removed with 100 mM NaOH prior to antibody staining (data not shown).

Data Base Searching. The OWL protein sequence data base (17) was searched by using BLAST (18) with peptide sequences obtained by high-energy CID analysis or Edman degradation. Masses obtained by MALDI were used by MOWSE (internet e-mail server version 5.1; mowse@dl.ac.uk) to search a peptide mass data base constructed from a theoretical trypsin digest of the entire OWL data base (9). Typical parameters employed were a 15% gel-derived protein-mass tolerance, a 2- or 3-Da peptide-mass tolerance, and a partial cleavage scoring factor of 0.4.

RESULTS AND DISCUSSION

Fig. 1 illustrates a representative 2D preparative gel containing human A375 melanoma proteins from whole cell lysates. The spots selected for mass spectrometric analysis were chosen because of their abundance, reproducibility, and relative isolation. Our strategy for the identification and characterization of these proteins is outlined in Fig. 2. Protein spots from several preparative gels were excised, and spots of identical mass pI were pooled before purification from the gel matrix, digestion with trypsin, and peptide separation by HPLC. The yield and number of peptides recovered from each spot were not always proportional to the number of gel plugs excised or the amount and size of protein present, suggesting differential digestion susceptibility, sample handling loss, and occasional protein comigration. For each spot, sub-pmol aliquots of HPLC fractions were analyzed by MALDI. The list of masses obtained from each protein digest served as the focal point for guiding subsequent experiments for identifying and characterizing primary structural features in the protein(s). The number of peptides recovered from digestion of a protein in a 2D PAGE spot varied and depended on protein size and the occasional comigration of other proteins. Experimentally determined masses were used with MOWSE data-base searching to match theoretical peptide masses and attempt to predict protein identities. HPLC fractions containing peptide masses below 2 kDa were subjected to high-energy CID for peptide sequencing. Protein identities were established by using the determined sequences to search the OWL protein sequence data base by using BLAST. Edman degradation was used to sequence peptides with masses greater than ≈2 kDa, since larger peptides ionize less efficiently by liquid secondary ion mass spectrometry.

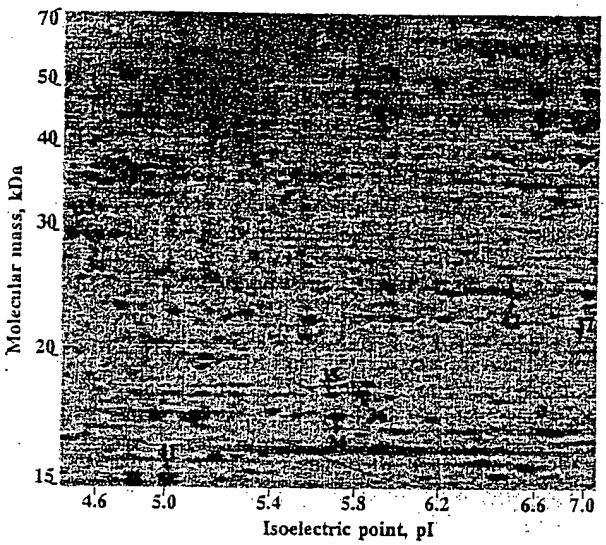


FIG. 1. Coomassie blue G-250-stained 2D preparative gel of human A375 melanoma proteins, numbered as in Tables 1 and 2. One milligram of total protein was loaded. The amount of protein estimated by densitometry in designated spots varied from 0.6 to 2.9 μg.

Tables 1 and 2 summarize peptide mass data, sequences determined or attributed by mass, and data-base search results for all spots studied. Ideally, we seek to attribute every mass to a unique sequence either by sequencing each peptide or matching it by mass to a peptide from the characterized protein.

From the 39 peptide masses from spot 3 listed in Table 1, MOWSE predicted the protein identity as α-enolase. Although most cell types demonstrate α-enolase activity, distribution of the specific enolase isoenzymes in biologically active dimers may be

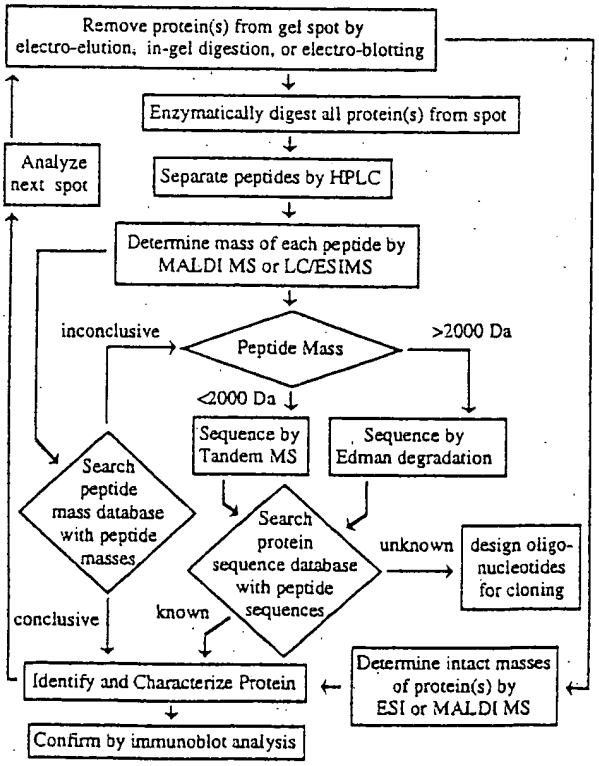


FIG. 2. Strategy for identifying and characterizing proteins separated by 2D PAGE. ESIMS, electrospray ionization mass spectrometry.

tissue specific. Only the α- and γ-homodimers can be detected in serum from patients with malignant ocular melanoma (19). During sequence analysis on selected peptides from spot 3, we also found a cytokeratin sequence LASYLDK that is present in many type I cytoskeletal keratins (mass/pI and peptide masses best matched cytokeratin 15). This indicated that at least two proteins had comigrated. Immunoblot analysis confirmed that α-enolase is present. Furthermore, the presence of a cytokeratin is consistent with the localization of many keratin subtypes to this gel region (20). However, the presence of cytokeratin was not predicted by MOWSE when masses exclusive of those attributable to α-enolase were used.

Protein disulfide isomerase (PDI) and stathmin were found in spots 30 and 24, respectively, and MOWSE readily predicted both. Sequencing of several peptides by high-energy CID established these identities. PDI, which catalyzes formation and interconversion of disulfide bonds in the endoplasmic reticulum, has been implicated in the activation of interferon-inducible genes in chronic myelogenous leukemia cells (21). Immunoblot analysis confirmed the presence of PDI in spot 30 and showed stathmin presence in spot 24 and in two nearby (more acidic) spots. Serine-phosphorylated forms of stathmin have been found in T-lymphocytes (22). The ability to suggest additional posttranslationally modified or genetically variant isoforms of a protein was one motivation for incorporating immunoblot analysis into our strategy.

In spot 33 fibroblast nonmuscle tropomyosin was identified following high-energy CID. Similarly, another form of tropomyosin, cytoskeletal type, was identified in spot 34. The two protein sequences are 76% identical. High-energy CID showed that the masses of the N-terminal peptides from both proteins differed by only 1 Da, due to an internal sequence difference

Table 1. Summary of data obtained for nine human melanoma proteins from eight 2D PAGE gel spots

Spot no.	Protein identified OWL accession no. [Mass/pI] (MOWSE search prediction)	MALDI mass, Da	Δ*, Da	Peptide sequence determined by high-energy CID†	MALDI mass, Da	Δ*, Da	Peptide sequence consistent with mass†
3	α-Enolase	1806.5	0.5	(R)AAVPSGASTGIYEALELR(D)‡	1633.5	1.6	(R)EAMRIGAEVYHNLK(N)
	P06733 [45 kDa/6.3]	1406.8	−0.8	(R)GNPTVEVDLFTSK(G)	1542.7	0.9	(K)VVIGMDVAASEFFR(S)
	(enoa_human.swiss)	2718.8	2385.9	2042.3 1929.6 1718.5 1643.8§	1424.9	−1.7	(R)YISPDQLADLYK(S)
		1604.9	1503.3	1477.3 1475.3 1436.6 1392.1§	1180.8	−1.6	(K)GVSKAVEHINK(T)
		1385.5	1325.5	1323.5 1268.6 1252.5 1236.1§	1141.5	−2.8	(R)IGAEVYHNLK(N)
		1112.7	1095.5	1026.8 984.0 954.7§	905.2	0.2	(R)IEEELGSK(A)
					899.4	−0.7	(K)TIAPALVSK(K)
					705.7	0.8	(K)GVPLYR(H)
	Cytokeratin 15 (type I)	810.4	0.5	(R)LASYLDK(V)	1799.1	0.0	(R)VLAEMREQYEAMAEK(N)
	P19012 [45 kDa/6.3]	1558.7	1044.8¶		1437.6	−1.0	(K)EVASNTEMIQTSK(T)
	(not predicted)‖				1099.2	−2.5	(R)EQYEAMAEK(N)
24	Stathmin	912.6	−0.4	(K)SHEAEVLK(Q)	1390.4	0.8	(R)ASGQAFELILSPR(S)
	P16949 [17 kDa/5.7]	888.7	−1.2	( )acASSDIQVK(E)	1330.2	2.7	(K)ESVPEFPLSPPK(K)
	(sthm_human.swiss)	1210.5§			1167.1	0.9	(K)AIEENNNFSK(M)
					1040.8	−0.4	(R)KSHEAEVLK(Q)
30	Protein disulfide isomerase	939.8	−0.2	(K)TVAYTEQK(M)	1583.8	3.0	(R)EATNPPVIQEEKPK(K)
	D16234 [54 kDa/5.8]	867.9	0.9	(K)YKELGEK(L)	1239.1	1.9	(R)DGEEAGAYDGPR(T)
	(humpicalf1.gb_pr)	790.0	2.1	(K)FLDAGHK(L)	1192.1	−0.2	(R)LAPEYEAATR(L)
		1459.1	1415.3§		998.2	2.0	(K)QAGPASVPLR(T)
33	Tropomyosin,	1241.1	−3.3	(R)IQLVEELDR(A)	1740.8	−3.1	(R)KIQALQQQADEAEDR(A)
	fibroblast non-muscle type	1198.8	−1.7	(M)acAGLNSLEAVKR(K)	1614.7	−1.0	(K)IQALQQQADEAEDR(A)
	P07226 [30 kDa/4.5]	721.0	−1.8	(R)AEFAER(S)	1297.5	−2.1	(R)KLVILEGELER(A)
	(tpmg_human.swiss)‖	1690.4	1492.7	1427.7 1410.5 1253.2 1192.6§	1272.1	−1.3	(R)EKAEGDVAALNR(R)
		1023.7	893.0	799.3 762.1 696.6 689.8§	1170.0	−2.3	(K)AEGDVAALNRR(I)
					1168.6	−2.8	(K)LVILEGELER(A)
					1014.7	−1.4	(K)AEGDVAALNR(R)
34	Tropomyosin,	1285.1	−0.4	(R)KLVIIEGDLER(T)	1773.1	1.2	(R)KIQVLQQQADDAEER(A)
	cytoskeletal type	1202.0	0.4	(M)acAGITTIEAVKR(K)	1726.7	−2.2	(R)IQLVEEELDRAQER(L)
	P12324 [29 kDa/4.6]	1155.2	−2.2	(K)LVIIEGDLER(T)	1670.2	−2.7	(K)LVIIEGDLERTEER(A)
	(tpmi_human.swiss)	1044.4	−1.0	(M)acAGITTIEAVK(R)	1471.4	−2.2	(R)EQAEAEVASLNRR(I)
		829.1	−1.9	(K)ILTDKLK(E)	1380.9	−0.7	(K)YEEEIKILTDK(L)
		774.3	−1.5	(R)AELAESR(C)	1241.2	−3.2	(R)IQLVEEELDR(A)
		743.1	−1.8	(R)LATALQK(L)	1130.4	−2.0	(K)MELQEIQLK(E)
		1958.6	1697.8	1481.0 1395.9 1295.9 1282.3§	921.5	0.5	(R)EMDEQIR(L)
		1255.1	1168.2	1148.2 1073.1 934.3 872.0§	873.7	−1.2	(R)EVEGERR(A)
35	Cu/Zn superoxide dismutase	822.8	−3.2	(R)TLVVHEK(A)	1023.5	−1.6	(K)HGGPKDEER(H)
	P00441 [18 kDa/5.7]	688.1	−1.7	(K)VWGSIK(G)	771.7	−2.1	(K)crESNGPVK(V)
	(not predicted)‖	2230.4	1972.6	1498.5 1151.8 1127.6 1101.2 1039.3§	662.8	0.1	(K)TGNAGSR(L)
36	Nucleoside diphosphate	1344.3	−1.3	(R)TFIAIKPDGVQR(G)	1193.1	−0.3	(K)crDRPFFAGLVK(Y)
	kinase A	1149.2	−1.2	(K)DRPFFAGLVK(Y)			
	P15531 [18 kDa/5.8]	1065.1	−1.1	(R)GDFCamIQVGR(N)			
	(ndka_human.swiss)‖	1800.7	825.6	656.4§			
41	Galaptin	1661.3	−1.5	(R)FNAHGDANTIVCamNSK(D)	1545.2	0.6	(K)crDSNNLCamLHFNPR(F)
	P09382 [16 kDa/5.1]	1502.2	0.6	(K)DSNNLCamLHFNPR(F)	1075.5	−1.5	(K)DGGAWGTEQR(E)
	(not predicted)‖	877.3	−0.8	(K)SFVLNLGK(D)	969.0	−0.1	(K)LPDGYEFK(F)
		2400.1	2210.1	2067.3 1991.4 1950.1 1940.4 1849.8§	909.5	−0.5	(K)crIKCamVAFD( )
		1822.0	1803.4	1781.1 1680.1 1528.3 1516.4§			

*The difference between the measured mass and calculated mass (average isotopic).
†Abbreviations: ( ), residues before/after peptide; Cam, acrylamide-modified Cys; Camo, Cam (oxidized); acX, acetylated N terminus; crX, carbamylated N terminus; Mso, Met sulfoxide.
‡Sequenced by Edman degradation.
§Masses neither identified nor attributed.
¶Masses ambiguous: α-enolase or cytokeratin.
‖Inconclusive search result.

Table 2. Summary of data obtained for triosephosphate isomerase from 2D PAGE gel spots 37 and 42

Spot no.	Protein identified OWL accession no. [Mass/pI] (MOWSE search prediction)	MALDI mass, Da	Δ*, Da	Peptide sequence determined by high-energy CID†	MALDI mass, Da	Δ*, Da	Peptide sequence consistent with mass†
37	Triosephosphate	1647.7	3.0	(K)crDCamGATWVVLGHSER(R)	3026.6	−4.8	(K)ELASQPDVDGFLVGGASLKPEFVDIINAK(Q)
	isomerase	1620.9	−2.0	(K)VTNGAFTGEISPGMIK(D)	2208.2	0.7	(K)VPADTEVVCamAPPTAYIDFAR(Q)
	P00938 [25	1615.6	−0.2	(R)RHVFGESDELIGQK(V)	1836.8	−2.4	(K)VAHALAEGLGVIACamoIGEK(L)
	kDa/7.0]	1599.2	−2.6	(K)DCamGATWVVLGHSER(R)	1820.6	−2.6	(K)VAHALAEGLGVIACamIGEK(L)
	(tpis_human.swiss)	1595.1	−8.8	(K)VVLAYEPVWAIGTGK(T)	1638.9	0.0	(K)VTNGAFTGEISPGMsoIK(V)‡
		1459.0	−0.6	(R)HVFGESDELIGQK(V)	1607.5	1.6	(K)LDPKIAVAAQNCamYK(V)
		1408.7	−6.9	(K)QSLGELIGTLNAAK(V)	1591.5	0.6	(K)VVFEQTKVIADNVK(D)
		1380.9	−3.7	(R)crIIYGGSVTGATCamK(E)	1503.8	1.2	(R)crHVFGESDELIGQK(V)
		1231.2	1.8	(K)DCamGATWVVLGH(S)	1468.7	1.1	(K)TATPQQAQEVHEK(L)
		1196.1	0.7	(K)crIAVAAQNCamYK(V)	1341.3	−0.3	(R)IIYGGSVTGATCamK(E)
		1168.7	0.3	(K)IAVAAQNCamoYK(V)	1278.2	2.7	(K)VIADNVKDWSK(V)
		956.4	1.3	(K)FFVGGNWK(M)	1234.4	−0.9	(K)SNVSDAVAQSTR(I)
		892.5	−1.5	(K)crVVFEQTK(V)	1151.2	−1.2	(K)IAVAAQNCamYK(V)
		758.9	0.0	(K)VIADNVK(D)	1123.1	−3.2	(R)crKFFVGGNWK(M)
		1747.1	1643.5	1481.4 1469.8 1109.4 826.5§	1081.9	−1.4	(R)KFFVGGNWK(M)
42	Triosephosphate	1760.1	2.1	(K)DCamGATWVVLGHSERR(H)	3035.3	3.9	(K)ELASQPDVDGFLVGGASLKPEFVDIINAK(Q)‡
	isomerase	1602.6	0.8	(K)DCamGATWVVLGHSER(R)	2342.9	6.2	(K)VAHALAEGLGVIACamIGEKLDER(E)‡
	P00938 [25	1407.7	−7.9	(K)QSLGELIGTLNAAK(V)	2203.9	−3.6	(K)VPADTEVVCamAPPTAYIDFAR(Q)‡
	kDa/6.4]	1340.1	−1.5	(R)IIYGGSVTGATCamK(E)	1826.4	3.2	(K)VAHALAEGLGVIACamIGEK(L)‡
	(tpis_human.swiss)	1235.7	0.4	(K)SNVSDAVAQSTR(I)	1615.6	−0.2	(R)RHVFGESDELIGQK(V)
		953.1	−2.0	(K)FFVGGNWK(M)	1604.7	0.8	(K)VVLAYEPVWAIGTGK(T)‡
		2210.7	2089.8	2034.2 1823.4 1594.4 1400.8§	1543.4	−0.4	(R)KQSLGELIGTLNAAK(V)
		1063.0	1060.4¶		1467.1	−0.5	(K)TATPQQAQEVHEK(L)
					1457.9	−1.7	(R)HVFGESDELIGQK(V)
					1151.5	−0.9	(K)IAVAAQNCamYK(V)
					1082.6	−0.7	(R)KFFVGGNWK(M)
					757.8	−1.1	(K)VIADNVK(D)
					747.0	−0.8	(R)EAGITEK(V)

*The difference between the measured mass and calculated mass (average isotopic).
†Abbreviations: ( ), residues before/after peptide; Cam, acrylamide-modified Cys; Camo, Cam (oxidized); acX, acetylated N terminus; crX, carbamylated N terminus; Mso, Met sulfoxide.
‡Sequenced by Edman degradation.
§Masses neither identified nor attributed.
¶Masses ambiguous: α-enolase or cytokeratin.
‖Inconclusive search result.

of ITTI (spot 34) vs. LNSL (spot 33). The presence of N-terminal acetylation is evident in Fig. 3 from the 42-Da increase in the mass of all N-terminally derived *a*- and *b*-type ions. Leucine at residues 3 and 6 is evident in Fig. 3 from the mass values of the *w*6 (656.3 Da) and *w*9 (970.4 Da) ions resulting from side-chain fragmentation. Isoleucine was likewise assigned at those same positions in the corresponding peptide from spot 34 (data not shown).

Differentiation and unambiguous structure assignment of these similar peptides illustrates the power of high-energy CID in protein primary structure determination. Methodology and fragment-ion nomenclature for high-energy CID are more thoroughly described elsewhere (12, 23). Although their biological function in nonmuscle cells is not clear, the tropomyosins are multiple isoforms which complex with microfilaments (24). While spot 34 was readily predicted as cytoskeletal tropomyosin by MOWSE, fibroblast nonmuscle tropomyosin was the only isoform among several ambiguous possibilities predicted for spot 33. Immunoblot analysis confirmed that tropomyosin is present in spots 33 and 34 and two more basic spots nearby.

Nucleoside diphosphate kinase A (NDPK-A) was identified in spot 36 following high-energy CID. NDPK catalyzes the phosphorylation of nucleoside 5′ diphosphates. NDPK-A is the product of the *nm23* gene, which, when overexpressed, decreases tumorigenesis of melanoma cell lines (25). MOWSE results for this spot were inconclusive, due in part to chemical modification of two peptides.

Cu/Zn superoxide dismutase (Cu/Zn SOD) and galaptin were identified in spots 35 and 41, respectively, following high-energy CID. Immunoblot analysis confirmed the presence of both Cu/Zn SOD and galaptin. Cu/Zn SOD catalyzes the conversion of toxic superoxide to hydrogen peroxide. Galaptin, a lectin present in many tissues, has been found in several skin tumor types including melanoma, and its reduced expression level has been suggested as a means of diagnosing malignancy (26). MOWSE was unable to predict the identity of both of these proteins from the obtained MALDI data. Only 15–25% of the MALDI masses obtained for spots 35 and 41 were consistent with unmodified tryptic peptides. Our studies show that for any given MOWSE search, numerous proteins present in the data base appear to be randomly capable of matching ≈30% of the masses in a given list. Below this level we have little confidence in MOWSE predictions and find sequence determination necessary for protein identification.

Triosephosphate isomerase (TPI), present in both spot 37 and spot 42, was readily predicted by MOWSE (Table 2). The success of MOWSE is noteworthy despite the fact that several of the MALDI masses are quite inaccurate (an error of more than ±3.5 Da, attributed to an inconsistently performing laser power supply). Subsequent high-energy CID and Edman degradation elucidated the sequences of several covalently modified peptides and established the identity of TPI in both spots. Peptides that represented a nonspecific trypsin cleavage between His-95 and Ser-96 and chemical modifications, including acrylamide-modified cysteines, oxidation, and carbamylation, were found. The chemical modification of cysteine arose from protein interaction with the gel matrix. Carbamylation is likely a consequence of using 2 M urea buffer during enzymatic digestion of spot 37. In-gel digestion with 100 mM ammonium bicarbonate was done for spot 42 rather than electroelution, thus eliminating the need for urea.

TPI, which catalyzes the interconversion of dihydroxy acetone phosphate and glyceraldehyde 3-phosphate in glycolysis, gluconeogenesis, fatty acid synthesis, and the pentose shunt, is

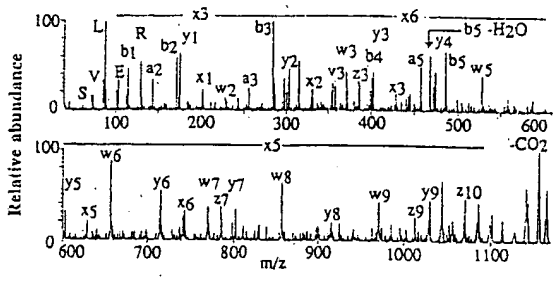


FIG. 3. High-energy CID spectrum of acetylated N-terminal tryptic peptide of nonmuscle tropomyosin (1–10 pmol) from spot 33, MH^+ = 1199.7 (monoisotopic mass). Sequence: CH_3CO-AGLNSLEAVKR. Peptide backbone cleavage ions associated with charge retention at the C terminus are denoted by *x*, *y*, and *z*; and at the N terminus by *a* and *b*. Side-chain fragment ions are denoted by *v* and *w*.

the product of a single gene. However, multiple electrophoretic forms have been observed, and some result from deamidations at Asn-71 and Asn-15 or oxidation of Cys-126 (27). By high-energy CID, an Asn-71 containing peptide in spot 37 (mass 1620.9 in Table 2) was not deamidated. The corresponding peptide was not found in spot 42. All five cysteines in TPI were shown by high-energy CID, Edman degradation, or mass matching to be consistent with acrylamide modification for spots 37 and 42. Throughout this work we have found cysteines exclusively in the acrylamide-modified form, despite inclusion of the antioxidant sodium thioglycolate during electrophoresis.

We have not currently pursued the strategy outlined in Fig. 2 which uses mass spectrometry to determine accurate masses on intact proteins isolated by 2D PAGE. For proteins in solution, mass determinations accurate to 0.01% (±2 Da on a 20-kDa protein) may be obtainable by MALDI or electrospray ionization mass spectrometry (28). Accurate intact masses could suggest the existence and possible identity of posttranslational modifications or confirm an expected protein sequence. However, mass determination on intact proteins isolated by 2D PAGE is confounded by problems associated with removal of stained protein from the gel matrix in a format amenable to mass spectrometry. Initial attempts at MALDI from proteins and peptides blotted onto polyvinylidene difluoride and nylon membranes are encouraging (29–31).

Since our strategy is based on obtaining sequence information at fmol to low pmol levels, discovery of previously unknown peptide sequences will enable design of oligonucleotide probes for gene cloning. Furthermore, the growth of sequence data bases fueled by the Human Genome Project will increasingly reduce the chances of discovering unknown proteins. However, the discriminating power of a peptide mass-matching strategy employing ±2–3 Da MALDI mass data will progressively decline with data-base growth, thus producing increasing numbers of inconclusive results, necessitating sequencing. Enhanced mass accuracy obtained by incorporating electrospray ionization mass spectrometry or improvements in MALDI technology would slow the anticipated decline in discriminating power of mass-matching strategies by reducing the effective portion of the data base considered in a search. Hence, when protein recognition is sought without need for primary structure characterization, the identities of many known proteins may be rapidly predicted by combining mass matching with direct MALDI analysis of unfractionated digests (5, 6, 9, 10). Unfortunately, the ability to directly perform substantial sequence analysis is severely compromised. Of the 11 proteins we characterized from 10 gel spots, 6 were readily predicted by mass matching with MOWSE, and 9 contained unmatched covalently modified peptides. Consequently, our results illustrate the necessity of peptide sequencing and the advantages of tandem mass spectrometry to rapidly and unambiguously identify proteins isolated by 2D PAGE.

We thank F. C. Walls for technical expertise, and Fisons Instruments for loan of a VG TofSpec. We are grateful to Dr. S. Hanash for stathmin monoclonal antibody and to Drs. L. Pohl, L. Oberly, J. Zieske, and H. Allen and Chemicon International for antisera. This work was supported by National Institutes of Health Grants DK26743, RR01614, and ES04705 to A.L.B. and Grants CA27903 and CA44446 to L.B.E.; National Science Foundation Grant DIR-8700766 to A.L.B.; University of California Tobacco Research and Disease Related Program Grant 2RT0089; and a University of California Systemwide Biotechnology Grant.

1. O'Farrell, P. H. (1975) *J. Biol. Chem.* 250, 4007–4021.
2. Smith, D. M., Tran, H. M. & Epstein, L. B. (1995) in *Cytokines: A Practical Approach*, ed. Balkwill, F. (Oxford Univ. Press, London), in press.
3. Celis, J. E., Rasmussen, H. H., Olsen, E., Madsen, P., Leffers, H., *et al.* (1993) *Electrophoresis* 14, 1091–1198.
4. Hall, S. C., Smith, D. M., Masiarz, F. R., Soo, V. W., Tran, H. M., Epstein, L. B. & Burlingame, A. L. (1993) *Proc. Natl. Acad. Sci. USA* 90, 1927–1931.
5. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C. & Watanabe, C. (1993) *Proc. Natl. Acad. Sci. USA* 90, 5011–5015.
6. Rasmussen, H. H., Mortz, E., Mann, M., Roepstorff, P. & Celis, J. E. (1994) *Electrophoresis* 15, 406–416.
7. Toussi, A., Paquet, N., Huber, O., Frutiger, S., Tissot, J., Hughes, G. & Hochstrasser, D. (1992) *J. Chromatogr.* 582, 87–92.
8. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. (1987) *Proc. Natl. Acad. Sci. USA* 84, 6970–6974.
9. Pappin, D. J., Hojrup, P. & Bleasby, A. J. (1993) *Curr. Biol.* 3, 327–332.
10. Ji, H., Whitehead, R. H., Reid, G. E., Moritz, R. L., Ward, L. D. & Simpson, R. J. (1994) *Electrophoresis* 15, 391–405.
11. Walls, F. C., Hall, S. C., Medzihradszky, K. F., Yu, Z., Burlingame, A. L., Evans, S., Hoffman, A. D., Buchanan, R. & Glover, S. (1993) *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics* (Am. Soc. Mass Spectrom., San Francisco), pp. 937a–937b.
12. Medzihradszky, K. F. & Burlingame, A. L. (1994) *Methods Companion Methods Enzymol.* 6, 284–303.
13. Mann, M., Hojrup, P. & Roepstorff, P. (1993) *Biol. Mass Spectrom.* 22, 338–345.
14. Patton, W. F., Pluskal, M. G., Skea, W. M., Buecker, J. L., Lopez, M. F., Zimmermann, R., Belanger, L. M. & Hatch, P. D. (1990) *BioTechniques* 8, 518–527.
15. Rosenfeld, J., Capdevielle, J., Guillemot, J. C. & Ferrara, P. (1992) *Anal. Biochem.* 203, 173–179.
16. Hines, W. M., Falick, A. M., Burlingame, A. L. & Gibson, B. W. (1992) *J. Am. Soc. Mass Spectrom.* 3, 326–336.
17. Bleasby, A. J. & Wootton, J. C. (1990) *Protein Eng.* 3, 153–159.
18. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403–410.
19. Shine, B. S., Hungerford, J., Vaghela, B. & Sheraidah, G. A. (1990) *Br. J. Ophthalmol.* 74, 427–430.
20. Sun, T. T., Tseng, S. C., Huang, A. J., Cooper, D., Schermer, A., Lynch, M. H., Weiss, R. & Eichner, R. (1985) *Ann. N.Y. Acad. Sci.* 455, 307–329.
21. Johnson, E., Henzel, W. & Deisseroth, A. (1992) *J. Biol. Chem.* 267, 14412–14417.
22. Strahler, J. R., Hailat, N., Lamb, B. J., Rogers, K. P., Underhill, J. A., Melhem, R. F., Keim, D. R., Zhu, X., Kuick, R. D., Fox, D. A. & Hanash, S. M. (1992) *J. Immunol.* 149, 1191–1198.
23. Biemann, K. (1990) *Methods Enzymol.* 193, 455–479.
24. Lees-Miller, J. & Helfman, D. (1991) *BioEssays* 13, 429–437.
25. Leone, A., Flatow, U., King, C. R., Sandeen, M. A., Margulies, I. M., Liotta, L. A. & Steeg, P. S. (1991) *Cell* 65, 25–35.
26. Santa-Lucia, P., Wilson, B. D. & Allen, H. J. (1991) *J. Dermatol. Surg. Oncol.* 17, 653–655.
27. Gracy, R. W., Yuksel, K. U., Chapman, M. L. & Dimitrijevich, S. D. (1990) *Prog. Clin. Biol. Res.* 344, 787–817.
28. Chait, B. T. & Kent, S. B. (1992) *Science* 257, 1885–1894.
29. Strupat, K., Karas, M. & Hillenkamp, F. (1994) *Anal. Chem.* 66, 464–470.
30. Vestling, M. M. & Fenselau, C. (1994) *Anal. Chem.* 66, 471–477.
31. Zaluzec, E. J., Gage, D. A., Allison, J. & Watson, J. T. (1994) *J. Am. Soc. Mass Spectrom.* 5, 230–237.

Document: Declaration Pursuant to 37 C.F.R. §1.132
Attorney Client-Matter No: 66654-669

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"EXPRESS MAIL" MAILING LABEL NUMBER: EV 462932419 US

DATE OF DEPOSIT: May 26, 2004

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 CFR 1.10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO COMMISSIONER FOR PATENTS, P.O. BOX 1450, ALEXANDRIA, VA 22313-1450.

Leand Bontales
Printed Name of Person Mailing Paper or Fee

Signature of Person Mailing Paper or Fee

PATENT
Client-Matter No.: 66654-669
(P-LJ 4859)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of	)	Confirmation No: 7275
Ruoslahti and Pasqualini	)	Group Art Unit: 1632
Serial No.: 09/922,227	)	Examiner: S. Priebe
Filed: August 2, 2001	)	
For: METHODS OF IDENTIFYING MOLECULES THAT HOME TO A SELECTED ORGAN *IN VIVO*	)	

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION PURSUANT TO 37 C.F.R. § 1.132

Sir:

I, Erkki Ruoslahti, declare as follows:

1) I am the Erkki Ruoslahti who is named as a co-inventor of the above-identified patent application.

2) I understand that the claims of the subject application stand rejected, in part, on the basis that one skilled in the art allegedly would not have been able to identify homing molecules by *in vivo* panning with untagged libraries of molecules at the time the priority application was filed.

3) I believe that in 1995, at the time the priority application for the above-identified application was filed, an ordinary scientist using the teachings of the specification would have been able to use untagged libraries such as peptide and small molecule libraries in the claimed *in vivo* panning methods to recover and identify molecules that home to a selected organ or tissue.

4) Corroboration of identification of homing molecules using an untagged small molecule library in accordance with the teachings of the patent specification is provided herein in paragraphs 5 to 8, which describe identification of a homing molecule from a library of 10 small molecules injected into the circulation of mice. As described further below, mass spectrometric analysis of extracts from organs harvested 10 minutes following intravenous injection of the library showed that one compound homed selectively to brain. These results will shortly be published in the peer-reviewed journal of ChemBiochem. The galley proofs of this publication (Brown et al., ChemBiochem. 5:1-9 (2004)) are attached as Exhibit A.

5) The library of 10 small molecules was selected from a much larger library containing pharmacologically active compounds. The structures of the 10 small molecules, which were selected by an individual who had no involvement with the subsequent *in vivo* panning experiments, and their molecular weights are shown in Table 1. Each of the ten compounds was present at a final concentration of 1 mM in phosphate buffer (40 mM, pH 7.2).

6) To identify homing molecules, two-month-old female BALB/c mice were anesthetized with avertin. Mice were injected intravenously in the tail-vein with 200 µl of library solution (200 nmol per compound). After ten minutes of circulation, organs were harvested and washed with phosphate-buffered saline (PBS). Each organ was mixed with 5 ml acetone and homogenized with a Handishear hand-held homogenizer (Virtis; Gardener, NY). Organ/acetone homogenates were transferred to 15 ml centrifuge tubes and incubated at -80°C for 12 hours to precipitate proteins. Following centrifugation at 3,000 x *g* for 30 minutes at 4°C, supernatants were recovered and dried in a SpeedVac. Organ extracts prepared from mice injected with PBS served as an internal control for the experiment.

7) Dried organ extracts were resuspended in 50 µl acetonitrile. The samples were then diluted 1:2 in 20 mg/ml alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 0.1% trifluoroacetic acid. Finally, one-half of the diluted sample was analyzed on an Applied Biosystems MALDI-TOF Voyager DE-Pro mass spectrometer at The Burnham Institute.

8) As shown in Figure 1, a striking peak with m/z = 301 was present in the brain extract from library-injected mice but was absent from brain extract of control mice. Based on the molecular weight of 301 Da, the brain-homing compound could be identified from the other components of the injected library as compound "2C11" (see Table 1). This 301 Da compound is a pharmacologically active benzodiazepine, Oxazepam, known to cross the blood-brain barrier. Furthermore, as shown in Figure 1, a biologically inactive 265 Da benzodiazepine was not detected in the brain (left panel, arrow beneath X-axis). It is striking that we were able to make this observation in a blinded manner and identify the benzodiazepine, even though the MALDI-TOF mass spectrometer was not optimally accurate at the molecular weight range in this study. These results indicate that organ-homing molecules can be isolated from a chemical library by *in vivo* panning and subsequently identified using mass spectrometry.

9) Further corroboration of the identification of homing molecules using an untagged small molecule library in accordance with the teachings of the patent specification is provided herein in paragraphs 10 to 20, which describe identification of homing molecules from a library of 75 random small molecules. As described in the paragraphs which follow, selected organs (kidney, liver, lung and brain) were harvested in organic solvent to precipitate proteins, and molecules from the library were subsequently identified in the soluble phase using electrospray mass spectrometry.

10) In particular, 75 organic compounds were randomly selected from a 420,000-member library from ChemBridge (San Diego, CA). There was high structural diversity among the 75 organic compounds, and the masses of the compounds differed from each other by at least 4 Da. The library was resuspended in dimethylsulfoxide (DMSO), with each individual compound at a final concentration of 1.33 mM. The 75 ChemBridge compounds and

their masses are shown in Table 2. *In vivo* panning was performed as described above, except that mice were injected intravenously in the tail-vein with 25 µl of library (33 nmols per compound), and control mice were injected with a single compound known to lack homing activity. In some cases, 250 pmol to 2.5 nmol of control compound (ChemBridge 5116670, molar mass of 340 Da) was added as a reference for quantification of the amount of homing molecule in a target organ. In addition, the organ extracts from mice injected with the 75-member library were not analyzed by MALDI-TOF mass spectrometry as described above but were prepared for subsequent analysis on a electrospray mass spectrometer. The dried organ extracts were resuspended in 100 µl of methanol, vortexed for about 10-20 minutes, and centrifuged to pellet debris. The supernatants were recovered, further diluted 1:20 in methanol, and 20 µl of the diluted sample was analyzed on a Waters® Micromass® LCT mass spectrometer (Milford, MA) at The Scripps Research Institute (La Jolla, CA). The mobile phase of the liquid chromatography was run with 90% methanol, 9% water and 1% acetonitrile. This form of mass spectrometry was more suitable than MALDI-TOF for analyzing compounds having a mass of 300-600 Daltons in tissue extracts.

11) The mass spectra of the tissue extracts from mice injected with the 75-compound library were analyzed as follows. Candidate organ-homing molecules were represented as various peaks which appeared in organ extracts from library-injected mice but were absent from the corresponding organ extracts of control mice. High intensity peaks that appeared in both extracts from library-injected mice and control mice were excluded from the analysis; using this criteria, compounds represented by peaks at m/z = 437 and m/z = 453 were excluded as potential organ homing molecules. Other spectral peaks from tissue extracts of library-injected mice were also found in extracts of control-injected mice when the sensitivity of detection was increased by focusing the mass spectrometer on a narrower mass range. Extraneous peaks observed only in extracts from control mice, including the oscillatory signals observed in some analyses, also were excluded. In addition, compounds represented by peaks present at significant amounts in several different organs, such as compounds having m/z = 408 and m/z = 424, were excluded as potential organ homing molecules due to a lack of specificity.

Peaks of interest were analyzed by comparison to the theoretical masses of the 75 individual ChemBridge compounds shown in Table 2. Figures 2A (kidney), 2B (liver), 2C (lung) and 2D (brain) show the results of initial screening experiments in which the library (upper panel) or control compound (lower panel) was injected into mice. Homing data from two different mice were generally analyzed. Where results between the two mice differed significantly, results from the second mouse are shown as an inset to the figure.

12) Representative spectra from a kidney extract prepared from a library-injected mouse and from a control kidney extract from a mouse injected with a non-targeting compound are shown in Figure 2A. High intensity peaks present in kidney extract from a library-injected mouse had masses as follows: 298, 392, 408, 424, 432, 437, 453, 500 and 518 (see Figure 2A). Of these, peaks at 437 and 453 were excluded as present in control extracts, and peaks at 408 and 424 were excluded as present in multiple different organ extracts. The remaining peaks (at 298, 392, 432, 500 and 518) were identified as representing potential kidney-homing compounds and were identified by comparison of m/z value of each peak to the theoretical mass of the 75 compounds in the starting library, as summarized in Table 3 below.

13) When tested individually for their ability to home to selected organs, ChemBridge compounds 5862461 and 6074428 accumulated in kidney and did not localize to any other tissue (Figures 3A and B, respectively). Compound 5343617 also localized to the kidney as shown in Figure 3C, although this compound accumulated preferentially in the liver and also was observed in the lung. Of the two remaining potential kidney-homing molecules, compound 5116670 accumulated non-specifically when injected individually while compound 5536652 was not tested. In sum, these results confirm that mass spectrometry can be used to identify organ-homing molecules such as kidney-homing molecules from an untagged small molecule library.

Table 3

Potential kidney-homing compounds identified in library screening

ChemBridge compound	Observed peak	Theoretical molecular weight	Homing activity observed when molecule injected individually
5343617	298	300.08	Predominantly liver; also lung and kidney
5116670	392	392.01	Non-specific accumulation in all four organs
5862461	432	432.07	kidney
6074428	500	500.01	kidney
5536652	518	520.01	not tested individually

14) Representative spectra from a liver extract prepared from a library-injected mouse (upper panel) and a control liver extract from a mouse injected with a non-targeting compound (lower panel) are shown in Figure 2B. In the liver extract from the library-injected mouse, nine high intensity peaks were observed (m/z = 298, 364, 392, 408, 424, 437, 453, 472 and 516). As before, peaks at 437 and 453 were excluded as present in control extracts, and peaks at 408 and 424 were excluded as common to several organ extracts. As summarized in Table 4, the remaining five peaks, with m/z = 298, 364, 392, 472 and 516, were identified as representing potential liver-homing compounds, and were subsequently injected into animals individually for further characterization of homing activity.

15) When tested individually for their ability to home to selected organs, ChemBridge compound 5343617 accumulated in liver, and to a lesser extent in lung and kidney,

corroborating that this molecule is a liver-homing molecule. The remaining molecules either accumulated non-specifically or were not detected when analyzed individually.

Table 4

Potential liver-homing compounds identified in library screening

ChemBridge compound	Observed peak	Theoretical molecular weight	Homing activity observed when molecule injected individually
5343617	301/302	300.08	Predominantly liver; also lung and kidney
5216419	364	364.03	Not detected
5116670	392	392.01	Non-specific accumulation in all four organs
6873050	472	472.00	Non-specific accumulation in all four organs
7567423	516	516.01	Liver, kidney and brain

16) The same library was analyzed by *in vivo* panning for identification of lung-homing compounds. Representative spectra from a lung extract prepared from a library-injected mouse (upper panel) and a control lung extract from a mouse injected with a known, non-homing compound (lower panel) are shown in Figure 2C. Approximately seven high intensity peaks were observed in the lung extract from the library-injected mouse, at m/z = 301/302, 392, 408, 424, 437, 453 and 582. As before, peaks at 437 and 453 were discounted

since they appeared in control extracts, and peaks at 408 and 424 were discounted for their non-selective accumulation in different organ extracts. By comparison of the remaining three peaks (at m/z = 301/302, 392 and 582) to the theoretical masses of the components of the small molecule library, four compounds (ChemBridge 5343617, 5116670, 7609370 and 5710134) were identified as representing potential lung-homing compounds (Table 5). Each of these molecules were subsequently injected individually into mice in order to corroborate lung homing activity.

17) As discussed above, ChemBridge compound 5343617 localized, in part, to the lung. The remaining three compounds, 5116670, 7609370 and 5710134, when injected individually, were observed to accumulate non-specifically or were not detected (see Table 5).

Table 5

Potential lung-homing compounds identified in library screening

ChemBridge compound	**Observed peak**	**Theoretical molecular weight**	**Homing activity**
5343617	298	300.08	Predominantly liver; also lung and kidney
5116670	392	392.01	Non-specific accumulation in all four organs
7609370	582	580.68	Not detected
5710134	582	584.46	Lung, liver and kidney

18) The results of *in vivo* panning for identification of brain-homing molecules are shown in Figure 2D, which represents spectra from a brain extract prepared from a library-injected mouse (upper panel) and a control brain extract from a mouse injected with a control, non-targeting compound (lower panel). As shown in Figure 2D, no brain homing molecules were identified in the library of 75 random small molecules which was screened. The only high intensity peaks present in the brain extract from the library-injected mouse were also present in the mouse injected with control compound.

19) Representative individual homing experiments are shown in Figures 3A, 3B and 3C. Figure 3A shows detection of compound 5862461 (m/z = 429, 431 and 433) in a kidney extract after intravenous injection. Figure 3B shows that compound 6074428 (m/z = 499) localizes to the kidney following intravenous injection. Figure 3C shows that compound 5343617 (m/z = doublet at 298 and 300) homes to liver and lung following intravenous injection. These results, and the structures of the various ChemBridge molecules, are summarized in Figure 4.

20) Several of the organ-homing molecules contained one or more bromine atoms. Bromine exists in nature as two approximately equally abundant isotopes differing in mass by two daltons. When organ-homing molecules were assayed individually, molecules containing bromine exhibited characteristic split spectral peaks. In particular, ChemBridge molecule 5343617, which contains a single bromine atom, displayed a characteristic doublet at m/z = 298 and 300 (see Figure 3C). Similarly, ChemBridge molecule 5862461 which has two bromine atoms, displayed characteristic spectral peaks in the ratio of approximately 1:2:1 at m/z = 429, 431 and 433 (Figure 3A, inset). These characteristic spectral peaks corroborate identification of the bromine-containing ChemBridge molecules.

21) In sum, these results demonstrate that one of skill in the art would have been able to identify homing molecules following *in vivo* panning of untagged small molecule libraries using techniques such as mass spectrometry, which were routine in the art at the time the priority application was filed.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that any such willful false statement may jeopardize the validity of the application or any patent issued thereon.

Date: May 25 By: _____
Erkki Ruoslahti

Drug Identification through in vivo Screening of Chemical Libraries

Darren M. Brown, Maurizio Pellecchia, and Erkki Ruoslahti*[a]

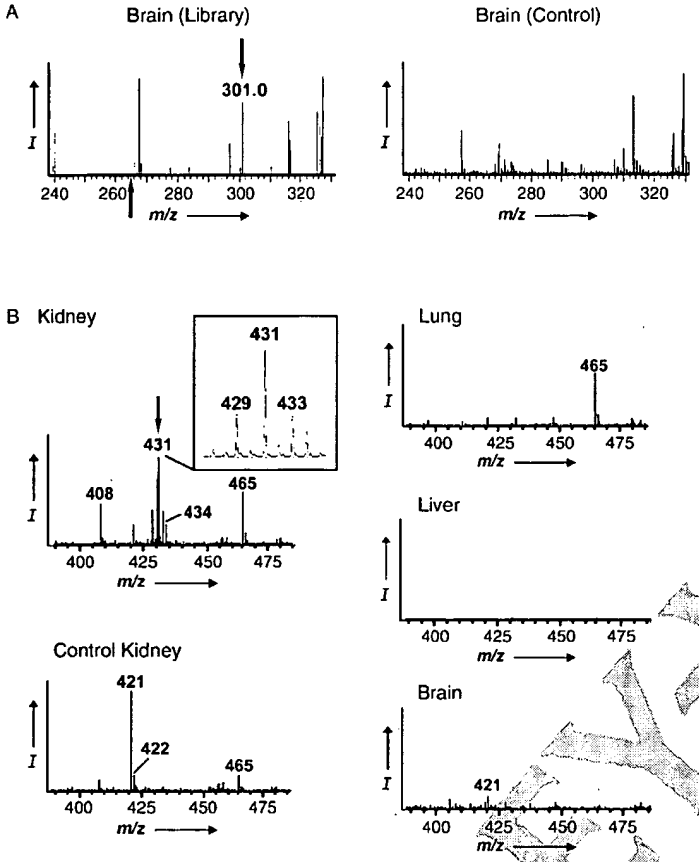
Specific cell-surface molecules can direct leukocytes and certain tumor cells to particular organs.[1-3] Recent work by our group has shown that peptides, selected by using the in vivo phage-screening approach, are also capable of mediating selective in vivo localization of phages to individual organs as well as tumors.[4-6] To develop our targeting technology beyond peptide-based systems, we investigated the feasibility of screening a chemical library to identify small molecules other than peptides that possess a preferential affinity for particular organs or tissues. As a proof of principle, we screened two different chemical libraries in vivo and identified three compounds that preferentially accumulated in individual organs: a pharmacologically active benzodiazepine localized in the brain, another compound specifically homed in on the liver, and the third on the kidneys. These results show that it is possible to use in vivo chemical-library screening to identify compounds that distribute themselves to specific sites in the body. Such knowledge can focus drug discovery on compounds with promising pharmacokinetic and tissue specificity profiles.

For in vivo screening of chemical libraries, we injected mixtures of small molecules into the circulation of mice, harvested selected organs in organic solvent to precipitate proteins, and detected the presence of compounds from the library in the soluble phase by mass spectrometry. We initially tested a library of ten compounds. Mass-spectrometric analysis of extracts from organs harvested 10 min after the intravenous injection of the library showed that one compound preferentially accumulated in the brain (Figure 1A). Breaking the code for the compounds revealed this 301 Da compound to be a benzodiazepine known as Oxazepam.[7] A biologically inactive 265 Da benzodiazepine also present in the library was not detected in the brain (Figure 1A). Thus, it appeared possible to obtain organ-targeting small molecules by screening chemical libraries in vivo. We also learned from these early studies that it was easier to detect library compounds and differentiate them from endogenous tissue molecules in organic extracts analyzed by electrospray mass spectrometry when the library molecules had molar masses greater than 300 Da.

To test a larger library, we assembled a mixture of 75 compounds with molar masses between 300 and 600 Da and screened for compounds that home in on the brain, liver, lungs, or kidneys. Mass spectrometry performed on organ extracts from library-injected mice identified ten molecules as candidate organ-homing compounds. These ten compounds

[a] *Dr. D. M. Brown, Prof. M. Pellecchia, Prof. E. Ruoslahti*
Cancer Research Center, The Burnham Institute
10901 North Torrey Pines Road, La Jolla, CA 92037 (USA)
Fax: (+1) 858-646-3198
E-mail: ruoslahti@burnham.org

 DOI: 10.1002/cbic.200300824



were tested individually for their ability to specifically target individual organs. Compounds 5862461 and 6074428 were found to accumulate in the kidneys (Figure 1B and C). The other tissues tested negative for these two compounds. Compound 5343617 was found primarily in the liver and, to a lesser extent, the lungs and kidneys (Figure 1D). The spectral patterns of compounds 5862461 and 5343617 were particularly distinct because these compounds contain bromine, which exists as two equally abundant natural isotopes,[8] and causes a characteristic two-mass-unit split in the spectral peak (Figure 1B, inset). One compound accumulated in the lungs, kidneys, and liver, but not the brain; and another localized to the brain, kidneys, and liver, but not the lungs (data not shown). These compounds are likely to bind to receptors that are expressed in more than one tissue, but the varying tissue selectivity of these compounds clearly indicates tissue-specific homing. Extracts from the organs of control-injected mice confirmed that no molecules matched the spectral pattern of the homing compounds. Two other candidate organ-homing compounds localized to all four tested organs. These compounds might bind to molecules present in all tissues, but it is also possible that their concentration in blood remaining in tissues is high enough to allow detection. As these compounds did not show any tissue-specific homing, we did not study them further. For three compounds, the specific organ homing could

Dateiname: Z300824
Seite: 2 te von 9
Datum: 15 KW., 8. April 2004 (Donnerstag)
Zeit: 14:33:05 Uhr
Pagina: 2
Umfang (Seiten): 9
Sprache American

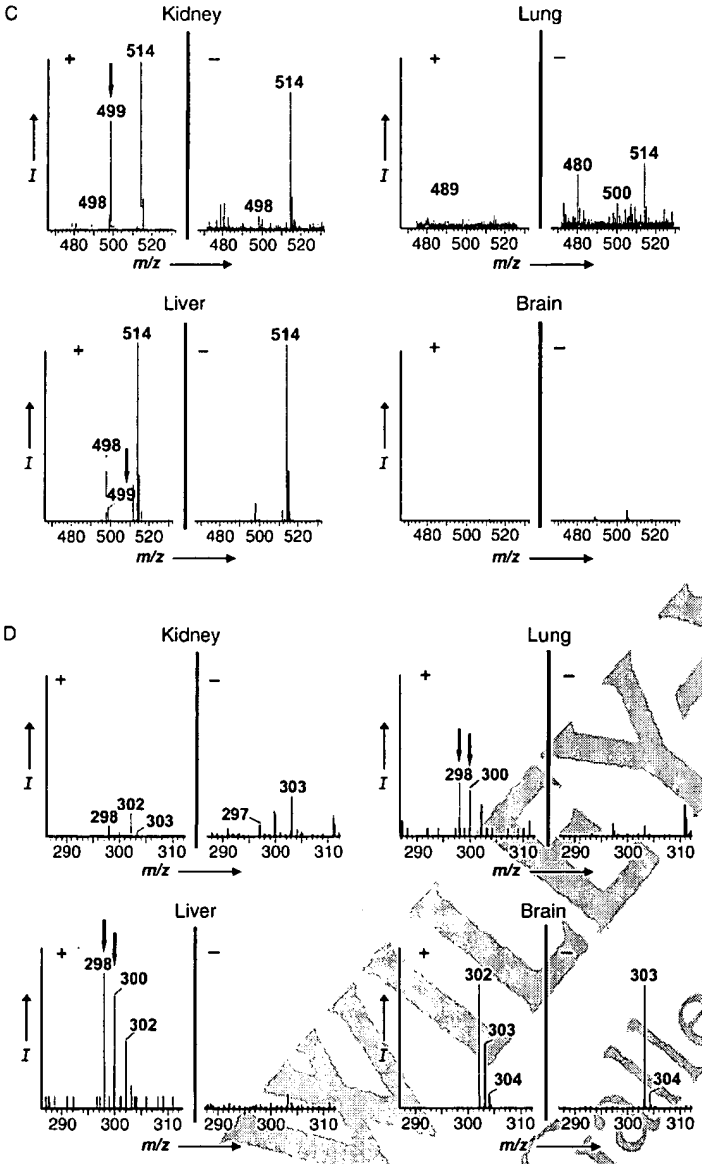


Figure 1. In vivo targeting of small molecules to particular organs. A) Detection of the benzodiazepine, Oxazepam, in the brain 10 min after intravenous injection with a ten-compound library. "Control" mice were injected with vehicle alone. The downward pointing arrow denotes the spectral peak for Oxazepam. The arrow below the axis denotes the m/z of the biologically inactive benzodiazepine in the library. No peak is seen at this position. B–D) Mice were intravenously injected with individual compounds from the 75-member library, and tissues were analyzed 10 min later by mass spectrometry. B) Detection of compound 5862461 in the kidneys after intravenous injection and circulation for 10 min. "Control" denotes mice injected just with DMSO. C) Compound 6074428 targets the kidneys ■■ or lungs?■■. D) Compound 5343617 targets the liver and lungs. Compound peak heights are shown as relative signal intensity (I). A "+" denotes compound-injected mice and "–" denotes DMSO-injected mice. The downward pointing arrows mark the spectral peaks for the organ-homing compounds.

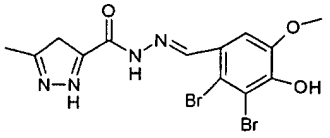
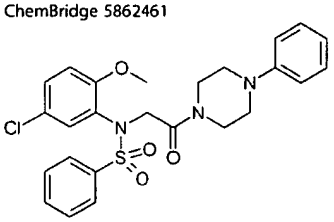
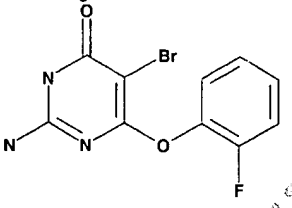
not be confirmed in individual testing. The remaining 68 compounds were not detected in any tissue, apparently because they did not sufficiently accumulate in any of the test tissues to bring the concentration above the detection limit.

We next quantified the organ accumulation of the three compounds with the most promising organ-homing proper-

Dateiname: Z300824
Seite: 3 in von 9
Datum: 15 KW., 8. April 2004 (Donnerstag)
Zeit: 14:33:05 Uhr
Pagina: 3
Umfang (Seiten): 9
Sprache American

ties. We used the mass spectrometer to compare the relative amounts of targeting compound in extracts of different organs. Compound 6074428 was at least 30-fold more concentrated in the lungs ■■or kidneys?■■ than in the liver, kidneys, and brain (Table 1). At least 2.4 times more compound

Table 1. Homing specificity of compounds and their accumulation in target organs. The structure, target organ, and homing activity of the three organ-homing compounds are shown. The quantity of homing compound in the target organs 10 min after an intravenous injection of individual compounds was determined as described in the Experimental Sectoin. The accumulation of targeting compound was expressed as normalized signal intensity level relative to the detection limit.

Organ-targeting Compound	Target Organ(s)	Signal Intensity[a] (fold higher than detection limit)
 ChemBridge 5862461	kidney	2.4-fold (±0.6)
 ChemBridge 6074428	■■lung?■■ liver lung	32-fold (±5.4) 67-fold (±2.1) 8.2-fold (±0.6)
 ChemBridge 5343617	kidney	1.2-fold (±0.3)

[a] Accumulation data is represented as mean percentage (±standard deviation) for two experiments per variable.

5862461 localized to the kidneys than to the liver, lungs, and brain. Compound 5343617 accumulated very strongly in the liver; about 55-fold higher levels were detected in the liver than in the kidneys, which contained a trace amount of the compound. This compound was also present at moderate levels in the lungs, but was not detectable in the brain. As each of these three compounds accumulated in different tissues ■■see above■■, their organ-selective homing is clearly specific and not due to the presence of blood or unspecific trapping in the target organs.

We then measured two parameters that influence the sensitivity of in vivo chemical-library screening. First, we used the mass spectrometer to analyze the spectral intensity of nine different compounds added to organ extracts, and found that the smallest amount of an individual compound that could be

detected in a tissue extract was between 34 and 215 pmol. For the second parameter, we determined the smallest amount of homing compound that could be injected and still detected in our in vivo screening system. For this analysis, the signal intensity of compound 6074428 in kidney extracts from mice injected with various amounts (2 to 125 nmol) of the compound was determined by mass spectrometry. The spectral peak at *m/z* 499 from compound 6074428 was detectable in kidney extracts from mice injected with as little as 7.8 nmol of targeting compound (Figure 2). In the initial library screen with 75

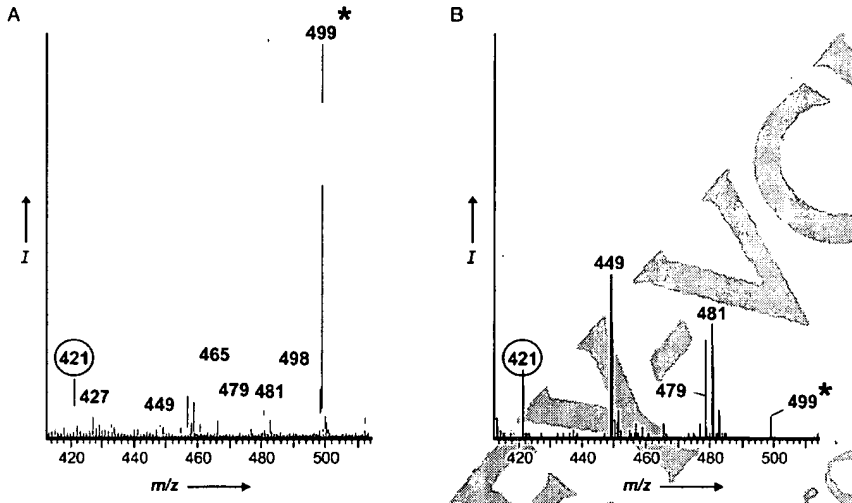


Figure 2. The lower detection limit of in vivo chemical-library screening. Mass-spectrometric analysis of kidney extracts from mice injected with either A) 125 nmol or B) 7.8 nmol of compound 6074428. The peak intensities were normalized to the height of an endogenous tissue molecule at m/z 421 that was consistently detected in kidney extracts (circled). The asterisk denotes the spectral peak for the kidney■■ ■■ homing compound, 6074428. The spectral peak intensities of other endogenous tissue molecules (e.g. the molecules at m/z 449 and 481) varied between experiments; as a result, they were not used to normalize the spectral peak intensities of compound 6074428. Compound peak heights are shown as relative signal intensity (I).

compounds, about 33 nmol of each molecule was present in the injected library mix. Therefore, it is likely that 300 compounds could be tested in a single screening round for organ-targeting compounds. Given the ease and simplicity of this screening technique, a library of 10000 compounds could be screened in vivo in a few weeks with a relatively small-scale effort.

We encountered some limitations with in vivo chemical-library screening that will be addressed in future studies. The volume of library injected into the mice (25 μL) was limited by the toxicity of the solvent, dimethyl sulfoxide (DMSO). With a less toxic solvent, it should be possible to inject up to 200 μL of library and screen potentially as many as 3600 compounds in one round. Emulsifying agents like Cremophor® EL, Emulphor®, polysorbate 80, Solutol® HS15, or solvents containing *N*-methylpyrrolidone could be used as an alternative to DMSO when solubilizing the chemical library before in vivo screening. In addition, only 1% of the organ extract could be analyzed by mass spectrometry due to the presence of various endogenous tissue compounds in the acetone extracts. A more selective ex-

traction and prepurification method could increase the sensitivity of the compound detection by mass spectrometry.

The biological basis for the targeting activity of some of the compounds identified in the screen has yet to be determined. However, it seems likely that binding to benzodiazepine receptors mediated the brain-homing activity of the pharmacologically active benzodiazepine, as the related inactive compound did not accumulate in the brain. The kidney-homing compound, 6074428, contains a benzenesulfonamide group that is known to have diuretic properties; perhaps this group mediates the kidney-homing activity of this compound.

This work provides the first demonstration that it is possible to conduct large-scale screening of chemical libraries in vivo. Such screening can identify targeted small molecules for use in a variety of applications and has some advantages over previous methods. In vivo phage screening primarily targets the vascular endothelium. Low-molecular-weight chemical compounds can target the vasculature, but are also likely to gain access to parenchymal cells in tissues. That parenchymal cells can be targets is suggested by our recovery of a benzodiazepine as a brain-homing molecule, as most receptors for these compounds are on the neurons. As an additional advantage, this screening approach does not require encoded or tagged library compounds. This is an improvement over other approaches that require separate chemistries for coupling different small molecules to synthetic or genetically engineered tags such as bacteriophages.[9] In addition, the absence of compound tags eliminates the possibility of interference by the tag with the in vivo homing activity.

The localization of selective molecules to specific "addresses" on the endothelium suggests that each tissue puts a specialized signature on its vasculature.[10] Organ-specific vascular molecules are attractive targets for the delivery of therapeutics to particular sites. By conjugating targeting moieties to drugs, diseases such as cancer can be treated with increased efficacy and fewer side effects;[11,12] phage-derived homing peptides and peptidomimetics have been used in this manner to target malignant tumors.[11–15] Organ-homing compounds isolated from chemical libraries are likely to be useful for similar purposes.

In vivo screening may also identify small molecules that have pharmacological effects at the target organ. The identification of a neuroactive compound and a potential diuretic as brain- and kidney-homing molecules, respectively, suggests that this may be possible. Thus, in vivo screening has the potential to advance drug discovery; it allows pharmacokinetics and specificity of action to be studied among large numbers of candidate compounds, or even from completely random libraries. Such approaches may accelerate the discovery and development of new drugs.

Experimental Section

A library of ten small molecules with molecular weights between 200 and 300 Da was prepared by a person not involved in the in vivo experimentation and was tested blindly. The ten-compound library was prepared in phosphate buffer (40 mM, pH 7.2) with each

molecule at a final concentration of 1 mM. A larger library of small molecules was prepared from 75 organic molecules (purchased from ChemBridge, San Diego, CA) with molecular weights between 300 and 600 Da. The library compounds were randomly selected from a 420000-member ChemBridge library, with each compound satisfying the following criteria: 1) the partition coefficient, expressed numerically as log*P*, was less than 5 and 2) the molecular weights of the compounds differed from each other by at least 4 Da. There was high structural diversity in the library, given that the only limitation was the compounds selected from the 420000-member parent library needed to fit the parameters described above. The 75-compound library was resuspended in DMSO, with each molecule at a final concentration of 1.33 mM.

To identify molecules that localize to particular organs, two-month-old female Balb/c mice were anesthetized with avertin (0.15 μL g^{-1}) administered intraperitoneally. In experiments with the ten-compound library, 200 μL of library solution (200 nmol per compound) was intravenously injected into the tail vein. With the 75-compound library, 25 μL of library solution (33 nmol per compound) was intravenously injected into the tail-vein. After 10 min of circulation, the lungs, liver, kidneys, and brain were removed. We found 5–15 min to be optimal for the screening of intravenously injected phage for homing to individual tissues and tumors,[16] and we wanted to keep the time short enough to prevent metabolism of the injected compounds, which would change their mass-spectrometric signature.

The organs were washed with PBS (5 mL) to remove excess blood and weighed. Each organ was mixed with acetone (5 mL) and then homogenized with a Handishear hand-held homogenizer (Virtis, Gardiner, NY). For certain organ homogenates, a control compound (ChemBridge 5116670, molar mass 340 Da, 0.25–2.5 nmol) was added as a reference to quantify the amount of homing compound in target organs. The organ/acetone homogenates were transferred to 15 mL centrifuge tubes and incubated at −80 °C for 12 h to precipitate the proteins. Following centrifugation for 30 min at 3000*g* and 4 °C, the supernatants were recovered and dried in a Speed Vac. A set of control organ extracts was also prepared from mice that were injected with pure DMSO (25 μL).

The dried organ extracts were resuspended in methanol (100 μL), spun in a vortex for about 10–20 min, and separated in a centrifuge to give pellet debris ■■ok?■■ The supernatants were recovered, further diluted 1:20 in methanol, and the diluted sample (20 μL) was analyzed on a Waters Micromass® LCT mass spectrometer (Milford, MA) at The Scripps Research Institute (La Jolla, CA). The mobile phase of the liquid chromatography was run with methanol/water/acetonitrile (90:9:1). By comparing the masses of the individual compounds and the molecules in the organ extracts of mice injected with DMSO to the molecules in the organ extracts from the mice injected with the library, we were able to identify molecules in the library that localized to a particular organ.

The accumulation of compounds in organs was measured as follows:

We first measured the signal intensities of the targeting compounds using the mass spectrometer and compared them to the signal intensity of a standard compound that was added to the organ extracts; this enabled us to normalize the intensity value of compound peaks from experiments performed on different days. We then determined the smallest amount of an individual compound that could be detected in a tissue extract using mass spectrometry by measuring the spectral intensity of nine different compounds added in small amounts to organ extracts. The detection

limit was defined as the spectral intensity level halfway between the background noise and the spectral intensity level generated from the smallest detectable amount of compound in organ extracts (averaged from nine different compounds whose spectra were analyzed and displayed with a scanning window of *m/z* 290–610). The normalized intensity values for homing compounds were compared to the detection limit to determine the degree of enrichment of compound in target organs relative to background levels. These enrichment values were not comparable from compound to compound, since each compound has a different ionization efficiency and stability on the mass spectrometer.

The Burnham Institute Animal Research Committee approved the animal experimentation in compliance wiht the relevent US laws.

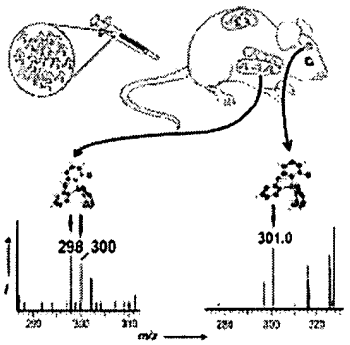
Acknowledgements

This work was supported by the U.S. National Cancer Institute (PO1 CA82713 and P30 CA30199) and the U.S. Department of Defense (DAMD17-02-1-0315). D.M.B. was supported by Post-doctoral training grant T32 CA09579 from the U.S. National Cancer Institute. We thank Dr. Eva Engvall for her valuable comments on the manuscript, and Roslind Varghese for editing.

Keywords: drug delivery · in vivo screening · mass spectrometry · tissue-specificity

[1] I. J. Fidler, *Surg. Oncol. Clin. North Am.* **2001**, *10*, 257.
[2] A. Müller, B. Homey, H. Soto, N. Ge, D. Catron, M. E. Buchanan, T. McClanahan, E. Murphy, W. Yuan, S. N. Wagner, J. L. Barrera, A. Mohar, E. Verastegui, A. Zlotnik, *Nature* **2001**, *410*, 50.
[3] E. J. Kunkel, E. C. Butcher, *Immunity* **2002**, *16*, 1.
[4] R. Pasqualini, E. Ruoslahti, *Nature* **1996**, *380*, 364.
[5] D. Rajotte, W. Arap, M. Hagedorn, E. Koivunen, R. Pasqualini, E. Ruoslahti, *J. Clin. Invest.* **1998**, *102*, 430.
[6] P. Laakkonen, K. Porkka, J. A. Hoffman, E. Ruoslahti, *Nat. Med. (N.Y. U.S.)* ■■ok?■■ **2002**, *8*, 751.
[7] ■■*title?*■■ 11th rev. ed. (Eds.: K. Z. Bezchlibnyk-Butler, J. J. Jeffries), Huber, Seattle, **2001**, p. 273.
[8] ■■*title?*■■ 84 ed. (Ed.: D. R. Lide), CRC Press, Boca Raton, **2003**, p. 11.
[9] T. F. Woiwode, J. E. Haggerty, R. Katz, M. A. Gallop, R. W. Barrett, W. J. Dower, S. E. Cwirla, *Chem. Biol.* **2003**, *10*, 847.
[10] E. Ruoslahti, *Nat. Rev. Cancer* **2002**, *2*, 83.
[11] W. Arap, W. Haedicke, M. Bernasconi, R. Kain, D. Rajotte, S. Krajewski, H. M. Ellerby, D. E. Bredesen, R. Pasqualini, E. Ruoslahti, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1527.
[12] F. Curnis, A. Sacchi, L. Borgna, F. Magni, A. Gasparri, A. Corti, *Nat. Biotechnol.* **2000**, *18*, 1185.
[13] J. D. Hood, M. Bednarski, R. Frausto, S. Guccione, R. A. Reisfeld, R. Xiang, D. A. Cheresh, *Science* **2002**, *296*, 2404.
[14] D. Hallahan, L. Geng, S. Qu, C. Scarfone, T. Giorgio, E. Donnelly, X. Gao, J. Clanton, *Cancer Cell* **2003**, *3*, 63.
[15] W. Arap, R. Pasqualini, E. Ruoslahti, *Science* **1998**, *279*, 377.
[16] J. A. Hoffman, P. Laakkonen, K. Porkka, M. Bernasconi, E. Ruoslahti in *Phage Display: A Practical Approach* (Eds.: T. Clackson, H. Lowman), Oxford University Press, Oxford, **2004**.

Received: November 19, 2003 [Z300824]



*D. M. Brown, M. Pellecchia, E. Ruoslahti**

■■-■■

Drug Identification through in vivo Screening of Chemical Libraries

Drug discovery process stood on its head? A proof-of-principle study has shown that large numbers of chemical compounds can be tested for selective accumulation in individual tissues in live mice. The picture shows how the site of accumulation of the tail-vein-injected molecules could be determined by mass spectrometry. The process picks drug candidates based on tissue specificity and pharmacokinetics rather than defined activity, which is the usual starting point of drug discovery.

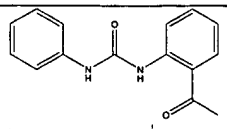
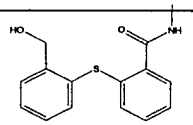
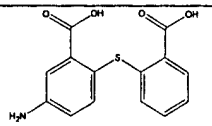
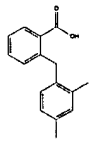
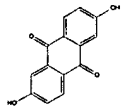
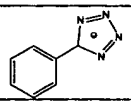
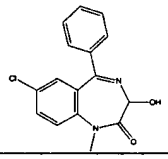
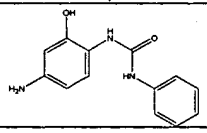
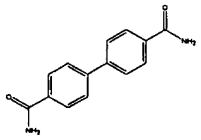
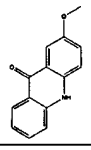
TABLE 1		
COMPOUND	STRUCTURE	MOLECULAR WEIGHT
1E4		226.2774
1B5		273.349
1B6		289.3052
1D6		240.3012
1B12		240.2148
2A6		264.3262
2C11		300.7439
2D8		243.2646
1B1		240.261
2E3		225.2464

Table 2

ChemBridge Compound number	Mass (g)
5343617	300.08
5135609	304.08
6098554	308.09
7246197	312.07
5904415	316.09
5108221	320.05
6955991	324
7477972	328.1
5225540	332.04
7253800	336.04
5231936	340.09
7383619	344.09
5403771	348.03
5279582	352.12
5377438	356.08
5550053	360
5216419	364.03
5276832	368.12
5155350	372.07
5809106	376
7257635	380.01
5225132	384.02
5380863	388.06
5116670	392.01
5624827	396.04

ChemBridge Compound number	Mass (g)
5578637	404.02
5217141	408.11
5300003	412
5326482	416.04
5246030	420.05
6090295	424.09
7384366	428.06
5862461	432.07
5364112	436.03
7100798	440.07
5569100	444.01
6903967	448.1
6170510	452.01
5169028	456.08
5214985	460.01
5216127	464
5255244	468.02
6873050	472
6987235	476.01
6872990	480
6875321	484.01
5130527	488.05
6987469	492.01
5348584	496
6074428	500.01

ChemBridge Compound number	Mass (g)
7583971	508.01
5768124	512
7567423	516.01
5536652	520.01
5717564	524.01
7497180	528.07
5671388	532
5670039	536.14
5555479	540.06
7575548	544.06
7591015	548.01
5374146	552.02
6394103	556.03
5557349	560.02
5551154	564.06
5711954	568.06
5101382	572.19
5227898	576.56
7609370	580.68
5710134	584.46
5751093	588.52
6968226	592.27
5743815	596.61
5233904	600.53